DOCUMENT RESUME

ED 413 364 TM 027 704

AUTHOR Cousins, J. Bradley; Donohue, John J.; Bloom, Gordon A.
TITLE Collaborative Evaluation: Survey of Practice in North America.
SPONS AGENCY Social Sciences and Humanities Research Council of Canada, Ottawa (Ontario).
PUB DATE 1995-11-00
NOTE 120p.; Paper presented at a Joint Meeting of the Canadian Evaluation Society and the American Evaluation Association (Vancouver, British Columbia, Canada, November 1995).
CONTRACT 410-92-0983
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Cooperation; Evaluation Methods; *Evaluators; Foreign Countries; *Program Evaluation; Program Implementation; *Social Science Research
IDENTIFIERS *Stakeholders

ABSTRACT

Forms of evaluation that involve evaluators working collaboratively with practitioners on applied social research projects are becoming increasingly common, but a body of empirical literature that warrants firm conclusions about collaborative evaluation has not yet been accumulated. This survey studies the views of evaluators and program practitioners using a conceptual framework developed to guide the study that considers pragmatic, philosophical, and political interests and three dimensions of the process: control of evaluation decision making, stakeholder selection, and depth of participation by program practitioners. An 8-page questionnaire was completed by 564 North American evaluators from professional association mailing lists. A subsample of 348 also selected and provided data on a recently completed collaborative evaluation. Sixty-seven practitioners who had collaborated returned parallel questionnaires. Findings show that evaluators tend to support pragmatic as opposed to political or philosophical justifications for collaborative evaluation and subscribe to a stakeholder-service orientation to the role. Attitudes toward such evaluation were generally positive, but they were found to depend on evaluators' experience with and involvement in such activities. Evaluators reported that practitioner involvement was generally more extensive than might be considered typical for traditional stakeholder-based evaluation, but was limited to mostly nontechnical research tasks. A wide range of stakeholder groups participated. Evaluators also tended to lean toward evaluator control rather than a balanced approach to evaluation decision making. In general, these self-reports found the impacts of the evaluation projects to be very favorable. Appendixes present the survey and cover letter. (Contains 86 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 413 364

Collaborative Evaluation: Survey of Practice in North America[1]

J. Bradley Cousins
John J. Donohue
Gordon A. Bloom

University of Ottawa

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. B. Cousins
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM027764

Monograph presented at the joint meeting of the Canadian Evaluation Society and the American Evaluation Association, Vancouver, November 1995.

---

[1] This research was funded by a grant from the Social Sciences and Humanities Research Council of Canada (# 410-92-0983). The opinions expressed are those of the authors and do not necessarily reflect Council policy. Correspondence: J. B. Cousins, Faculty of Education, University of Ottawa, 145 Jean-Jacques Lussier, Ottawa, ON, CANADA, K1N 6N5; bcousins@uottawa.ca. Note that the present version was revised from that delivered at the Evaluation '95 conference. This report has benefited from the helpful comments of Marvin Alkin and Nick Smith and the assistance of Stephanie Sutherland.

BEST COPY AVAILABLE

ABSTRACT

Forms of evaluation that involve evaluators working collaboratively with practitioners on applied social research projects are becoming increasingly common as indicated by the research and theoretical literature and practical reports. A body of empirical literature is accumulating but is currently not at a sufficient level of maturity to warrant firm conclusions about the nature, consequences and supporting conditions of collaborative evaluation. Moreover, while it is possible to determine from the literature what current theorists think about collaborative evaluation, little is known about the views of evaluators and program practitioners. The survey reported presently adds to knowledge about these issues and extends findings from mostly case study and small scale studies to a North American, continent-wide context.

A conceptual framework was developed to guide the study. Following Levin (1993), justifications or arguments for collaborative evaluation are categorized into three sets of interests: pragmatic (utilization, practical problem solving, organizational change); philosophical (technical quality of data, improved understanding, contribution to theory); and political (social justice, critical theory, emancipation). Also, three dimensions of form or process are proposed as being key to differentiating types of collaborative evaluation. They are control of evaluation decision making, stakeholder selection (who participates?), and depth of participation by program practitioners.

Five-hundred, sixty-four North American evaluators drawn from four professional association mailing lists returned an eight-page questionnaire. All completed a section of the questionnaire on views and opinions and background characteristics but a subsample of 348 also selected and provided data on a recently completed collaborative evaluation; some chose to nominate program practitioner/stakeholder colleagues to be surveyed about the same project. Sixty-seven such practitioners returned parallel questionnaires.

The research questions that guided the analyses were as follows:

For the complete sample of evaluators (N=564):

R1. What are evaluators' opinions and views about collaborative evaluation? How variable are these impressions?

R2. Do demographic / background variables differentiate evaluators' views and opinions about collaborative evaluation?

For the subsample of evaluators who selected a project on which to report (N=348):

R3. How do evaluators describe their collaborative evaluation projects? In what ways do they characterize its purposes and defining features? What types of impact and consequences do they report?

R4. To what extent are evaluators' reports about impact and consequences predicted by self-reported features of their projects? . . . their views and opinions about collaborative evaluation?

R5. Do demographic / background characteristics differentiate evaluators' descriptions and perceptions about the focal collaborative evaluation processes and consequences?

Finally, for the sample of practitioners who had been nominated by an evaluator colleague (N=67):

R6 To what extent do researcher and practitioner views and opinions about collaborative evaluation converge? Do perceptions about the process of the collaborative project and its impact converge? What are the nature of the differences in perception.

R7 Do demographic / background characteristics differentiate practitioner's descriptions and perceptions about the focal collaborative evaluation? . . . its processes and consequences?

The findings show that evaluators tend to support pragmatic as opposed to political or philosophical justifications, for collaborative evaluation and subscribe to as stakeholder-service orientation to the role. While attitudes toward such evaluation were generally positive, they were found to depend on evaluators' experience with and involvement in such activities. Evaluators reported that practitioner participation was generally more extensive than might be considered typical for traditional stakeholder-based evaluation but was limited to mostly non-technical research tasks. A wide range of stakeholder groups participated, especially primary stakeholders. Evaluators tended to lean toward evaluator control as opposed to a balanced approach to evaluation decision making. In general, the self-reported impact of the projects was found to be very favourable. Evaluations involving diverse groups of stakeholders had greater impact. Impact, dissemination patterns and collaborative evaluation processes were found to depend on evaluators' prior experience with the approach among other background characteristics. Generally, practitioner views and perceptions converged very well with those of the evaluators. While evaluators tended to be somewhat more optimistic about the evaluation's impact, program practitioners were more inclined that their participation caused them to reconsider their own practice in fundamental ways. Results are discussed in terms of the framework developed above. Limitations and caveats are noted. While the present data are unable to contribute to the professional debate about whether collaboration is a direction in which the field should be moving, they do add significantly to the empirical knowledge base concerning such practice.

TABLE OF CONTENTS

ABSTRACT I

TABLE OF CONTENTS iii

LIST OF TABLES AND FIGURES vi

CHAPTER 1: COLLABORATIVE EVALUATION IN PERSPECTIVE 1

1.1 INTRODUCTION AND BACKGROUND 1

1.2 CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW 2

1.2.1 Variations in Types of Collaborative Evaluation and Research 2

1.2.2 Conceptualizing Interest and Form in Collaborative Evaluation and Research 4

1.2.3 Research Based Knowledge about Participation and Collaboration in Evaluation 9

1.3 FOCUS FOR STUDY 11

CHAPTER 2: METHOD 13

2.1 GENERAL APPROACH 13

2.2 SAMPLE 13

2.3 INSTRUMENT 14

2.4 PROCEDURES 16

2.5 DESCRIPTION OF EVALUATOR RESPONSE 17

2.6 PLAN OF ANALYSIS 23

CHAPTER 3: EVALUATOR VIEWS AND OPINIONS ABOUT COLLABORATIVE EVALUATION **25**

3.1 DESCRIPTION OF RESPONSE PATTERNS 25

3.2 SCALE VARIABLE CONSTRUCTION AND DESCRIPTION 31

3.3 RELATIONSHIPS AMONG SCALE VARIABLES AND DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS 31

3.4 SUMMARY 34

CHAPTER 4: COLLABORATIVE EVALUATION PRACTICE **35**

4.1 DESCRIPTION OF RESPONSE PATTERNS 35

4.2 VARIABLE CONSTRUCTION AND DESCRIPTION 48

4.3 RELATIONSHIPS AMONG IMPACT, PROCESS AND OPINION VARIABLES 51

4.4 DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS 53

4.5 SUMMARY 54

CHAPTER 5: CONVERGENCE OF EVALUATOR AND PRACTITIONER PERSPECTIVES **56**

5.1 DESCRIPTIVE CHARACTERISTICS OF THE PRACTITIONER SAMPLE 56

5.2 VIEWS AND OPINIONS 56

5.3 COLLABORATIVE EVALUATION PRACTICE 58

5.4 TESTS FOR CONVERGENCE 62

5.5 EXPLAINING THE DIVERGENCE 66

5.6 DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS 72

5.7 SUMMARY 72

CHAPTER 6: DISCUSSION AND CONCLUSIONS **73**

6.1 LIMITATIONS 73

6.2 IMPLICATIONS FOR THEORY AND RESEARCH 74

6.2.1 Interests 74

6.2.2 Form 75

6.2.3 Other Findings 76

6.3 CONCLUDING REMARKS 77

REFERENCES **78**

APPENDICES **86**

APPENDIX A: Instrument and Data Summary 86

APPENDIX B: Letters of Informed Consent 94

LIST OF TABLES AND FIGURES

Tables

Table 2.1: Population and Sample Characteristics 15

Table 3.1: Opinions about General Issues in Evaluation 26

Table 3.2: Opinions about Practitioners' Participation in Formative Evaluation 28

Table 3.3: Scale Variable Construction--Views and Opinions 32

Table 3.4: Zero-order Intercorrelations among Opinion Variables--Evaluator Sample 33

Table 4.1: Frequency of Stakeholder Group Participation as Reported by Evaluators 37

Table 4.2: Actual Participation in Evaluation Activities 38

Table 4.3: Audiences for the Evaluation as Reported by Evaluators 42

Table 4.4: Modes of Communication of Recommendations as Reported by Evaluators 43

Table 4.5: Perceived Consequences of the Evaluation 47

Table 4.6: Variable Construction and Descriptive Statistics 50

Table 4.7: Zero-order Intercorrelations among Impact, Process and Opinion Variables 52

Table 4.8: Stepwise Multiple Regression of Impact on Process and Opinion 53

Table 5.1: Multivariate Analyses of Variance 63

Table 5.2: Written Comment Data from Discrepant Cases 68

Figures

Figure 1.1: Three dimensions of form in collaborative evaluation 7

Figure 1.2: Dimensions of form in collaborative evaluation represented in 3-D 8

Figure 2.1: Evaluator organizational location by response option selection 18

Figure 2.2: Evaluator years of experience by response option selected 19

Figure 2.3: Claim to evaluation as primary responsibility by response option selected 20

Figure 2.4: Percent evaluations considered collaborative by response option selected 21

Figure 2.5: Evaluator organizational location by nationality 22

Figure 4.1: Control of evaluation project decision making as reported by evaluators 40

Figure 4.2: Specification of conceptual framework as reported by evaluators 42

Figure 4.3: Perceived impact and intended uses of evaluation data 46

Figure 4.4: Perceived affective impact on evaluators 49

CHAPTER 1: COLLABORATIVE EVALUATION IN PERSPECTIVE

1.1 INTRODUCTION AND BACKGROUND

Collaborative evaluation is defined here as the application of principles of systematic inquiry to solve practical social or organizational problems. This general approach to applied research bears the distinguishing feature of evaluators (researchers) collaborating in some fashion with practitioners (non-researchers) in the production of applied research knowledge. Various manifestations of collaborative evaluation have enjoyed increased popularity and support among evaluation scholars and practitioners in recent years.

Shadish and Epstein (1987), in their survey of practising evaluators, found that one of four higher-order factors emerging from their data reflected a stakeholder-service pattern as a bona fide orientation toward the evaluation profession. More recently, Johnson (1995) observed that practising evaluators attribute greater degrees of utilization of data to increased participation in the evaluation process by practitioners. Several evaluation theorists (e.g., Cousins & Earl, 1992; Fetterman, 1994; Greene, 1988; Jenlink, 1994; Patton, 1994; Preskill, 1994a) have laid out compelling arguments favouring approaches to evaluation that involve in very direct ways those with a clearly identifiable stake in the object of evaluation (program, curriculum, policy, organization, etc.).

Despite growing interest in collaborative approaches to evaluation not all theorists agree about its legitimacy. Disagreements between Stufflebeam (1994) and Fetterman (1994, 1995) about the merits of empowerment evaluation provide a highly visible example. While it is possible to determine from the literature what some evaluation theorists think about various types of collaborative evaluation, little information is currently available about what practitioners of evaluation think of this approach and the extent to which they find it useful. The purpose of this study is to provide such information and to compare theses perceptions with those of the program practice/stakeholder community.

The present monograph adds to knowledge about collaborative evaluation in several ways. First, in the current chapter a conceptual framework is constructed and offered as an heuristic for locating and describing various forms of collaborative evaluation and other related species. Second, a 1995 large scale North American survey of evaluators is reported. That survey provides insights into (a) evaluators' opinions about a variety of issues concerning collaborative evaluation, (b) their reflections on a particular experience implementing collaborative evaluation within the past three years and (c) interrelationships among self-reported attitudinal variables, practical process factors and consequences of the evaluation. Finally, as part of the same survey, a subsample of evaluators nominated practitioners who collaborated on the specific evaluation identified, thereby permitting an analysis of the extent to which practitioner and evaluator self-reports converge.

1.2 CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

1.2.1 Variations in Types of Collaborative Evaluation and Research

In a recent study of participatory research in health promotion sponsored by the Royal Society of Canada (Green et al., 1995), a bibliography of published works on participatory research listed close to 500 titles. The list includes items dating back to Lewin (1946) with the vast majority surfacing within the last decade. With such a wide range of activity in the participatory research domain, some authors have taken it upon themselves to conceptually sort out the various forms of collaborative or participatory research that have appeared. Categorizations by King and Lonquist (1992), King (1995), Huberman (1995) and Garaway (1995) are particularly useful. Below, a variety of approaches are described citing representative authors.

Illuminatory evaluation: (Parlett & Dearden, 1977; Parlett & Hamilton, 1976). A social anthropological paradigm that advocates the use of flexibility in evaluation and relies on naturalistic and mixed-method approaches that would be sensitive to illuminating key program issues and the complex realities surrounding programs. The researcher maintains control of the process and works with practitioners in a consultatory mode.

Stakeholder-based evaluation: (Brandon, 1994; Bryk, 1983; Greene 1988; Mark & Shotland, 1985; Stake, 1976) This approach, emerging from Stake's (1976) notion of responsiveness in evaluation, advocates the involvement of a wide range of stakeholders in limited evaluation tasks, typically assisting in scoping out the evaluation and in interpreting findings. Often, stakeholder-based evaluation has been used in controversial program evaluation settings and has served to neutralized competing political forces, enhance utilization, or accommodate a wide range of interests. Mark and Shotland (1985) ground its rationale in three distinct areas: utilization, empowerment and representativeness in decision making.

Participatory action research (Whyte, 1991). The focus for participatory action research (PAR) as laid out by Whyte is the simultaneous improvement of practice and advancement of scientific knowledge through practitioner involvement in systematic inquiry. This approach is grounded in an organizational learning perspective in business, industry and agriculture. A similar approach called "action science" is put forward by Argyris and Schön (1991) and Argyris (1993). Note that the term participatory action research has been used to connote an ideologically distinct approach (e.g., McTaggart, 1991) that is described below under emancipatory action research. It may be noted that the term participatory has been used in a variety of ways in the literature. It is important to define the term since no common meaning seems evident.

Educative research: (Gitlin et al., 1992). Educative research values both experiential knowledge and knowledge produced through systematic inquiry and is decidedly normative aiming to

redefine the power base that exists in the present educational realm. That is to say, teachers direct participation in research will enable them through the power of knowledge.

Emancipatory and critical action research: (Carr & Kemmis, 1992; McTaggart, 1991a; Noffke, 1992; Tripp, 1990). This is a complex category of action research stemming from the initial writings of J. Habermas. It is characterized by a group of practitioners coming together with critical intent. Proponents advocate social change through enlightenment and action resulting from a process of deliberation and "symmetrical communication." Practitioners accept responsibility for "emancipation from the dictates of injustice, alienation and unfulfillment". Tripp suggests that in critical action research, critical reflection stops short of action, a distinction not found to be useful by McTaggart (1991) or Greene (1994).

Participatory evaluation (Brunner & Guzman, 1989; Cousins & Earl, 1992, 1995; Garaway, 1995; Greene, 1994; Shapiro, 1988). The approaches advocated by Brunner and Guzman (1989), Greene (1994) and Shapiro (1988) are decidedly normative and emancipatory in form and function and are characterized by participation in the evaluation of a wide range of stakeholder groups. The evaluator's role is one of facilitator (Brunner & Guzman, 1989). Through participation in the research and especially follow-up activities, social injustice is, in some sense, ameliorated. In contrast, the approach to participatory evaluation advocated by Cousins and Earl (1992, 1995) is grounded in pragmatic considerations of enhanced utilization of evaluation data and organizational problem solving capacity (see also Ayers, 1987; Lafleur, 1995; King, 1995).

Empowerment evaluation: (Fetterman, 1994, 1995). Rooted in community psychology and principles of self-determination, empowerment evaluation seeks to teach people to do their own evaluations and thus become more self-sufficient. Through facilitation, training and advocacy, evaluators foster illumination and liberation among program participants.

Collaborative action research: (Corey, 1953; Oja & Smulyan, 1989; Taba & Noel, 1957). Collaborative action research has its roots in education. Under the technical support of researchers, local actors undergo a process of reflection, investigation, action, interpretation and change. The central interest is their personal professional development and improved practice.

Teacher research: (Cochran-Smith & Lytle, 1993; Elliot, 1991; Hustler, Cassidy & Cuff 1986). Following work pioneered by L. Stenhouse, teacher research is action research conducted solely by teachers who, by virtue of their own tools of data collection and analysis (logs, observations, work products, reflective memos, annotated diaries, discussions) answer questions they themselves have posed.

School-based evaluation: (Alvik, 1995; Nevo, 1993). School evaluation carried out internally by school staff on an ongoing basis. Staff receive training and support from an external facilitator. Nevo (1994) advocates combining ongoing internal school-based evaluation with periodic

external (summative) evaluation. Data used to inform practice in a formative sense can be subsequently incorporated into an external summative review.

Democratic evaluation (MacDonald, 1976; McTaggart, 1991b) Democratic evaluation seeks to render evaluation decision making a democratic process, thereby enabling data to be used for legitimate agreed purposes. Key values inherent in democratic evaluation are pluralism, representativeness and "the right to know."

Developmental evaluation: (Patton, 1994). In developmental evaluation the process is the outcome, as the evaluator moves beyond evaluation responsibilities and plays an active role in program development. The approach is linked to the commonly used "improvement-oriented" connotation of formative evaluation, as opposed to formative evaluations originally intended usage as getting ready for summative evaluation (Scriven, 1991).

The forgoing typology, though loosely constructed, provides a glimpse of the variation in approaches to evaluation and applied social research that involve some level of collaboration between researchers and program practitioners/stakeholders. In the ensuing section, consideration is given to fundamental dimensions of purpose and the nature of collaborative evaluation. A conceptual framework useful for differentiating various types of systematic inquiry is constructed.

1.2.2 Conceptualizing Interest and Form in Collaborative Evaluation and Research

In thinking about any methodological entity it is always useful to consider purpose and to identify distinguishing descriptive features; the present case is no exception. In order to adequately capture the wide range of collaborative activities described above it seems useful to consider issues of purpose, goals, intention or interests as well as key differentiating dimensions of process or form.

Interests

Levin (1993), Mark and Shotland (1985) and Garaway (1995) each consider rationales for collaborative research activities. Levin's perspective seems particularly comprehensive and accommodating and will be used here. He suggests that rationales for collaborative work can be broken into three basic arguments or categories: the pragmatic, the philosophical and the political.

Pragmatic interests: The pragmatic argument is primarily concerned with the enhancement of intended use by intended users of evaluation data. Essentially, it is suggested that increased participation in research by stakeholders will heighten the probability that research data will have an intended impact. A growing body of data provide support for this perspective (Alkin, 1991; Alkin & Stecher, 1983; Bickel & Cooley, 1985; Bickel & Hattrup, 1995; Cousins & Earl, 1992, 1995; Cousins & Leithwood, 1993; Cousins & Simon, 1996; Greene, 1987; Huberman, 1990;

Johnson, 1995) generally defining impact as the instrumental (discrete decision is fostered) or conceptual (learning occurs) utilization of data.

More recently, pragmatic interests of collaborative evaluation have appeared to have expanded to consider consequences beyond the particular entity being evaluated. Cousins and Earl (1992) developed an organizational learning theoretical justification for their approach to participatory evaluation and subsequently compiled a set of empirical studies investigating effects on organizations of the participatory evaluation process (see also Cousins, 1995a). Similarly, other evaluation theorists have made cases for evaluators attenuating organizational consequences beyond program specific data utilization (Jenlink, 1994; Mathison, 1994; Patton, 1994; Preskill, 1994a, 1994b; Torres, 1994; Whyte, 1991). In one way or another, each of these arguments focuses on an expanded role for evaluators, one that embraces the developmental realities of organizations or programs as the case may be. As Jenlink envisions, "the evaluator, either internal or external, plays a pivotal role in working directly with the organization to examine the context, content, practices, and learning processes of the organization" (1994, p. 318).

While several of these approaches appear to mix evaluation roles with organizational development roles, this is not necessarily the case. In practical participatory evaluation, for example, the evaluator focuses on the development of skills in systematic inquiry among organization members such that the learning capacity of the organization is enhanced (Cousins & Earl, 1995). Since evaluation is inherently a social event, participation in carrying it out provides a forum for social interaction and exchange; sharing and dialogue about interpretations of ambiguous phenomena; the development of an understanding of cause and effect relationships of phenomena important to organization members. Such activity can, in theory, lead to incremental, first order changes in practice, or, depending on circumstances, to deeper, more fundamental rethinking of common assumptions and interpretations. Regardless, the evaluator`s main concern is the development of the capacity of organization members to do evaluation. Organizational consequences would follow such efforts.

Philosophic interests: A second set of interests that provides a rationale for participatory and collaborative research is epistemological in nature. The argument is that research knowledge and evaluation data are valid only when informed by practitioner perspectives (e.g., Guba & Lincoln, 1989). From this point of view Lewin's famous adage "nothing is so practical as a good theory" is reworked to highlight the contribution of good practice to the development of theory. Context as an essential component of validity is not embraced exclusively by those assuming a relativistic or interpretivist perspective. While traditional dissemination and utilization theories and models have been roundly criticised for their rigid positivistic assumptions and failure to deal with issues of context, recent "revisionist traditionalist" perspectives on D & U argue that contextual interpretation is an essential component of knowledge transfer and spread (Louis, 1995). The importance of participation, sustained interactivity, and linkages between knowledge production and utilization systems is a central feature of this perspective (Cousins & Simon, 1996; Huberman, 1994). As put by Huberman and Cox,

> It is the sustained interactivity that allows for the process of mutual education, by which users render evaluators progressively wiser in relation to the workings of the local context that actually account for the observed outcomes, and heighten the awareness of users by feeding back this information to users in ways that are locally recognizable yet framed in such a way as to deepen or sharpen local understandings (1990, p. 157).

Another dimension of the philosophic rationale is the level of abstraction at which produced knowledge will have an impact. For example, practical participatory evaluation (Cousins & Earl, 1995), traditional stakeholder-based evaluation (Bryk, 1983), utilization-focused evaluation (Patton, 1986) and other collaborative approaches to applied research are intended to have impact only at the level of practical problem solving and decision making. However, several forms of collaborative inquiry are deliberate mechanisms designed to produce valid social sciences knowledge or to influence theory (Argyris & Schön, 1991; Cochran-Smith & Lytle, 1993; Gitlin, et al., 1992; Whyte, 1991). Indeed the production of such knowledge is viewed by evaluators as an important dimension of the evaluation profession's contribution (Shadish & Epstein, 1987).

Political interests: Levin's third argument is grounded in political principles. Evaluation is inherently a political process as noted scholars remind us (Stufflebeam, 1994; Weiss, 1991a). But others advocate collaborative forms of research and evaluation as a way to enact political agendas. Feminist, emancipatory and critical action research approaches (Carr & Kemmis, 1992; McTaggart, 1991a), ideologically tied forms of participatory evaluation (Bruner & Guzman, 1992; Fetterman, 1994, 1995; Greene, 1994; Shapiro, 1988), and other forms of evaluation intended to bring about social justice (House, 1991) are examples. The argument is that "researchers have a moral obligation not just to study, but to act in the interests of those they study" (Levin, 1993, p. 332). Forms of participatory action research are geared toward simultaneously changing both individuals and the culture of groups, institutions and society (McTaggart, 1991a). House (1991) points out that competing conceptions of social justice (utilitarian, pluralist, critical) coupled with the acknowledgement that evaluators cannot be value neutral poses significant problems not just for evaluation but of evaluation for social justice.

Form

Cousins and Earl (1992) cast participatory evaluation as an extension of the traditional stakeholder-based model and proposed three basic distinguishing features: stakeholder selection, control of the evaluation and depth of participation. Participatory evaluation is intended to limit stakeholder participation to primary users or those with a vital interest in the program (Alkin, 1991). It is also intended to provide for partnership between researchers and practitioners and shared control over evaluation project decision making. Finally, stakeholders participate in all phases of the research including data collection, analysis, interpretation and reporting. Shulha and Cousins (1995) observed that these distinguishing features provide basic dimensions or continua along which any given collaborative research project might be located.

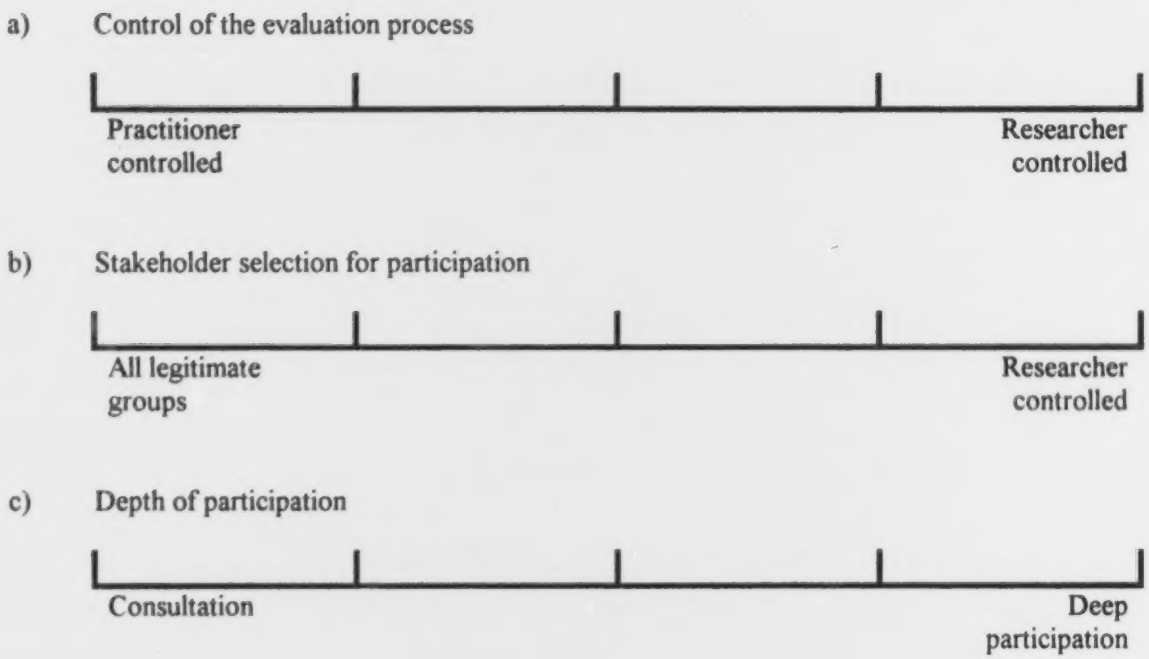


Figure 1.1: Three dimensions of form in collaborative evaluation.

Consider, participatory evaluation as described above in relation to Figure 1.1. It would be located in the middle of dimension (a), the right-hand side of dimension (b) and the right hand side of dimension (c). Similarly, traditional stakeholder based evaluation would be located on the left side of dimension (b) since a wide range of legitimate stakeholders are typically invited in (Mark & Shotland, 1985). Since the evaluator remains in control of the evaluation, and involves stakeholders in scoping out the evaluation and the interpretation of findings, it would also be located on the right-hand side of dimension (a) and somewhere in the middle of dimension (c). Participatory action research, as described by McTaggart (1991a) provides a third example. McTaggart suggests that in PAR teachers, for example, generally limit participation to themselves or a small group of colleagues, have complete control of the research process only relying on researchers for advice and consultation, and carry out all of the research tasks. PAR, as defined by McTaggart then, would be located on the left-hand, right-hand, and right-hand sides of dimensions (a), (b) and (c), respectively.

It seems reasonable, then, to conceive of these dimensions of form as being mutually independent or orthogonal (Shulha & Cousins, 1995). As such, they may be represented in three dimensional space as is done in Figure 1.2. Note that the quadrants defined by dimensions (a) and (b) are numbered 1 through 4 and that the plane set by those dimensions separates forward (f) and

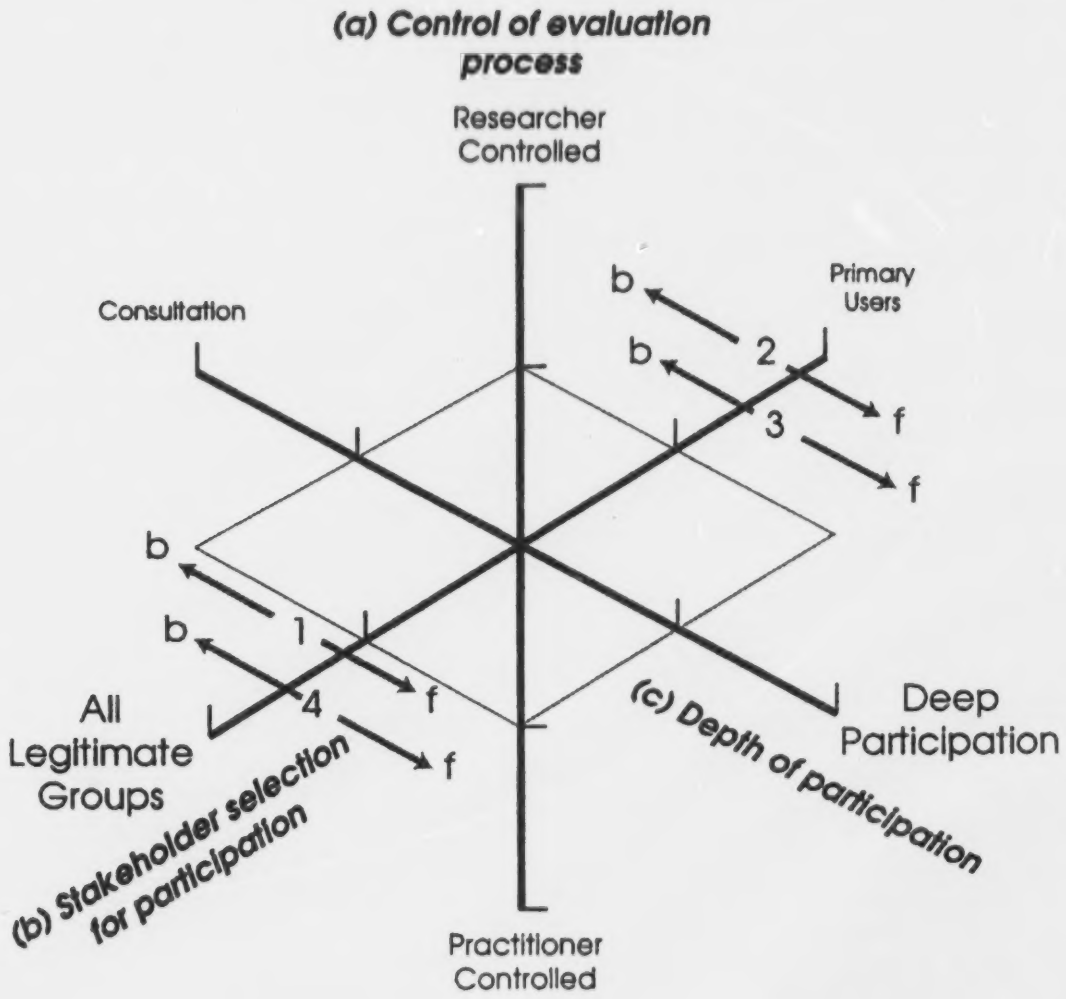


Figure 1.2: Dimensions of form in collaborative evaluation represented in 3-D.

back (b) domains. We would plot the three examples sited above as follows: Participatory evaluation (sectors 2f-3f); stakeholder-based evaluation (sectors 1b-1f); and PAR (sector 2f). The figure might also serve to usefully differentiate non-collaborative forms of evaluation such as Stufflebeam's (1994) objectivist evaluation where stakeholders are consulted for information and are involved as sources of data as opposed to participants in the process. Depending on the range of stakeholders consulted objectivist evaluation would be located in quadrant 1b-2b.

1.2.3 Research-Based Knowledge about Participation and Collaboration in Evaluation

While the literature is replete with descriptions of approaches to and advocacy for collaborative evaluation (e.g., Bragg, 1995; Cousins & Earl, 1992; Fetterman, 1994; Mathison, 1994; Preskill, 1994a; Owen & Lambert, 1994) there is a limited but growing body of research on the nature, consequences and conditions supporting its various forms. To follow is a brief summary of the main findings that have emerged from this literature.

Participant gratification: Several studies found that participation in research, though demanding in light of other ongoing commitments and responsibilities, was a very useful professional development experience. Such development was usually connected to research skill development but in some cases pertained to the refinement of understanding about program phenomena (Cousins, 1995a, 1995b, 1996; Earl, 1995). Bickell and Hattrup (1995) reported that effective collaboration means breaking out of traditional roles. While this may represent significant challenges for practitioners, in a survey study Cousins and Walker (1995) showed that teachers' attitudes toward participating in applied research were more favourable if they had done so previously. Related to this finding is a sense of "empowerment" experienced by some participants either at the individual level (due, for example, to promotion within the organization, Cousins, 1995a) or collectively as in the study by Mertens, Berkely and Lopez (1995) who showed that teachers in a culture that has not traditionally valued their opinion were willing to voice it after having participated in evaluation.

Stakeholder selection: The selection of stakeholders for participation appears to be a complex issue in need of further study. King (1995) found that how people were selected (volunteer vs. assigned) and whether they had time to participate mattered greatly to the quality of the experience. In a single case study of school-based evaluation Alvik (1995) observed that teachers who were initially resistant toward participating in the process changed their views after witnessing the benefits colleagues appeared to be deriving from it. Cousins (1995b) and Lafleur (1995) showed that if so called primary users, people with the organizational power to act on evaluation data, were not members of the research team, disastrous results in terms of utilization and follow-up were observed. Smith (1980, 1983), showed that trustees with limits on their time and low perceived need for involvement in school accreditation found invovlement in it to be less than desirable. There "appears to be little relationship between level of involvement and level of

utility for school board members, many of whom do not think it would be useful to be more involved in accreditation evaluations" (1980, p. 57).

Incentives and supports: It is very necessary for local organizations in which collaborative evaluation is taking place to both support and provide incentives for participation. Bickell and Hattrup (1995) indicate that collaborative research projects tend to be underestimated by participants in terms of time commitments. Cousins (1996) found that administrators often underestimate the time and effort required to carry out local applied research. This mistake can have severe consequences for those charged with the responsibility because the research may simply become responsibility added on to an already full workload. Shulha and Wilson (1995) provided a dramatic example of how quickly a collaborative enterprise can come crashing down once administrative support is removed. Alvik (1995) showed that methodological sophistication in schools is lacking and needs to be developed in schools. He, along with Cousins and Walker (1995) provided rich data on the lack of fit of research in the daily lives of teachers. This fit will need to be recognized and addressed if collaborative models of evaluation are to flourish in practice-based organizations.

Depth of participation: Complex, technical research tasks ought to be subcontracted or undertaken by researchers. The cost-benefit ratio of training non-researchers to handle these activities does not warrant the effort. On the other hand, involvement of primary users in data collection, interpretation and reporting was found to lead to utilization, professional development and gratification (Cousins, 1995b; 1996). Alkin and Stecher (1983) found that local involvement in school-based evaluation predicted utilization. They suggested that "the capability for performing evaluation must be shifted to the local schools" (p. 30). In his study of school-based evaluation in Israel, Nevo (1994) reached a similar conclusion concerning the conduct of formative evaluation. Levin (1993) report that practitioner involvement was limited to validating data. He found a general lack of appreciation for and utility of the research for participating members. Brandon (1994) involved program beneficiaries at the scope setting and interpretation phases of an evaluation. He reported that their participation helped to improve the validity of data, while at the same time enabling a relatively efficient research process.

Political influences: Several authors (Cousins, 1995b; King, 1995; Lafleur, 1995) found that politics play just as important a role in collaborative evaluation as in other forms. These studies reported mischievous use of findings and general lack of attention to important data by those in an organizational position to do something with them. Brickell and Hattrup (1995) underscore the importance of embracing conflict as part of the process and being willing to learn from it. They also noted significant cultural differences between researchers and teachers and the concomitant difficulties that arose.

Cultural change takes time: Some studies have found that collaborative evaluation has precipitated change in infrastructure or had other impacts on organizations above and beyond the evaluation project per se (Cousins, 1995a, 1995b, 1996; Earl, 1995), but it is apparent that

expectations for change of this sort ought to accommodate reasonably lengthy time frames. It seems likely that evaluation of this type will be well on its way to institutionalization before substantial developments in organizational learning capacity will be observed.

Unit of analysis: With an expanded role of evaluators and broader, more diffuse expectations for evaluation impact, there is a need to consider closely the appropriate unit of analysis. Cousins (1995b) showed different effects at different levels within a school system. The various units were, the school-based participatory evaluation team, the school, other schools in the district and the district administration.

Shared control: Many of the studies reported have seen the researcher playing a lead role in controlling the evaluation and research activities. Some evidence shows that where control is balanced higher level of engagement and ultimately utilization may be expected (Brickell & Hattrup, 1995; Cousins, 1995b). Cousins (1996) found that researcher visibility can intrude in interesting ways. He found that in a case where researcher involvement was protracted over a long period of time, false expectations were created and this worked to diminish the impact of the study. On the other hand, in cases where the researcher participated "behind the scenes" research supports were provided and the impact of the studies was substantial. Levin (1993) reported a multiple case study where control was dominated by researchers. He found that practitioners became somewhat defensive about criticism because researchers were essentially outsiders who came off looking as though they understood more than organization members.

Effects on researchers: The expansion in roles in order to accommodate collaborative evaluation is not limited to practitioners. Clearly, there is a significant demand on researchers for leadership, negotiating skills and instructional techniques. Tolerance for a slower pace than that to which one is normally accustomed is another important consideration (Cousins, 1996). As Levin (1993) put it, "the degree of advocacy for collaboration, however, is not nearly equalled by work which is actually collaborative. There is a simple and powerful reason why this should be true: writing about or advocating collaboration is much easier that doing collaborative research." (p. 332). Another important finding is that often times researchers find themselves being asked to be all things to all people. Levin (1993) showed that organization members eventually revealed they wanted a consultant, not a researcher. They viewed the project as the researchers' enterprise. Similarly, in a school-university partnership addressing the collaborative development of innovative approaches to classroom assessment, ultimately, teachers were disappointed that shiny new products had not been delivered by the researchers. Working as the evaluation coordinator for a series of funded school improvement projects, Lee found that she was also being called upon as a change agent (Lee & Cousins, 1995).

1.3 FOCUS FOR STUDY

The foregoing review provides many insights into the nature, consequences and supporting conditions for collaborative research. Many of these studies are small scale case

studies reporting either participant observation or interview data from practitioners involved in the collaborative enterprises. The authors are unaware of any large scale survey studies of evaluators working in this domain. Such a study would help to test and to generalize some of the preliminary findings reported above.

The present study reports findings from a large sample of evaluators, and as mentioned at the outset, a subsample of program practitioners working on collaborative evaluation projects. The robustness of many of the findings reported above are of interest but the intention is also to add new insights. The questions driving the study are as follows:

For the total sample.

R1. What are evaluators' opinions and views about collaborative evaluation? How variable are these impressions?

R2. Do demographic / background variables differentiate evaluators' views and opinions about collaborative evaluation?

For the subsample of evaluators reporting on a collaborative evaluation completed within the last three years.

R3. How do evaluators describe their collaborative evaluation projects? In what ways do they characterize its purposes and defining features? What types of impact and consequences do they report?

R4. To what extent are evaluators' reports about impact and consequences predicted by self-reported features of their project? . . . their views and opinions about collaborative evaluation?

R5. Do demographic / background characteristics differentiate evaluators' descriptions and perceptions about the focal collaborative evaluation process and consequences?

For the subsample of evaluators and program practitioners who reported on the same collaborative evaluation project.

R6. To what extent do researcher and practitioner views and opinions about collaborative evaluation converge? Do perceptions about the process of the collaborative evaluation project and its impact converge? What are the nature of differences in perception?

R7. Do demographic / background characteristics differentiate practitioners` descriptions and perceptions about the focal collaborative evaluation processes and consequences?

CHAPTER 2: METHOD

2.1 GENERAL APPROACH

The general design of the study is a cross-sectional survey of evaluators residing and working in North America. The survey questionnaire was based on the conceptual framework specified in chapter 1 with a primary focus on evaluation utilization (pragmatic) interests. Analyses were generally guided by the framework, although responses to some open-ended survey questions were content analysed for emergent patterns.

A unique aspect of the survey is that program practitioners/stakeholder collaborating with evaluators on evaluation projects were also invited to respond. These respondents were nominated by evaluators and subsequently sent the survey package. Undoubtedly, the achieved sample of practitioners undoubtedly is systematically biased toward those having positive experiences with collaborative evaluation as a consequence of (a) evaluators' decisions to nominate and the complete freedom they were given to select a nominee and (b) program practitioners' decision to participate. Nonetheless, this design feature presents an opportunity to compare the extent to which evaluator and practitioner perceptions compare. Given sample biases, points of non-convergence would be interesting to examine closely. The authors are unaware of other surveys of evaluators and researchers using the "paired sample" approach employed here.

Chapter 2 lays out the basic information about sample instrumentation, and procedures. Descriptive information is also presented to enhance the reader's knowledge of the achieved sample composition. Data analysis procedures that were used are then described.

2.2 SAMPLE

The achieved sample for the study was comprised of 564 evaluators and 67 practitioners. The response rate is difficult to estimate reliably for reasons that will become clear in the description of sampling procedures provided below.

While the views and opinions of a representative sample of evaluators in North America was desirable, the primary intention was to obtain a representative sample of evaluators who do collaborative evaluation. This was problematic since Shadish and Epstien (1987) observed four distinct patterns of evaluators among members of the American Evaluation Association (AEA) perhaps only one of which (stakeholder-service orientation) pertains to evaluators practising collaborative evaluation. Permission was sought and received to use membership mailing lists from four North American professional associations to which evaluators belong: (1) AEA (2) the Canadian Evaluation Society (CES), (3) Division H (School Evaluation) of the American Educational Research Association (AERA) and (4) the Association of Educational Research Officers of Ontario (AERO). Resource limitations precluded the possibility of obtaining a

representative sample of members from each of the organizations, leading to a decision to focus most heavily on AEA and CES as primary interdisciplinary evaluation associations. Also, resource limitations meant that only English-speaking evaluators could participate in the survey.[1] An interest was maintained in AERO and AERA since substantial collaborative evaluation activity is taking place in the educational sector (see, e.g., Cousins & Earl, 1995). AERO, a small Canadian association, is predominantly comprised of practising school district evaluators. It was decided, then, to limit the AERA population only to those members who were based in school districts (as evidenced by their mailing address).

Table 2.1 provides population and sample characteristics. Sampling was done on a stratified random basis with association representing the only stratum used. Almost 90% of the selected association members belonged to either AEA or CES representing significant proportions of these populations. The raw response rate of 28.2%, while not encouraging, must be interpreted in view of the survey focus. Several communications were received from evaluators who declined the opportunity to participate on the basis of their non-involvement in collaborative evaluation.[2] Of the 564 evaluators responding, 216 (38.2%) declined the opportunity to complete Part A, pertaining to a recently completed collaborative evaluation. These respondents completed Part B, views and opinions about evaluation, and Part C, background information, only. Of the 348 evaluators reporting on a specific collaborative evaluation, 116 (33.3%) chose to nominate a practitioner to be surveyed. Of the 116 surveys forwarded to practitioners, 67 (58.6%) usable returns were received.

2.3 INSTRUMENT

An eight-page survey questionnaire was developed and checked closely for continuity and clarity of phrasing. This instrument and a summary of descriptive statistics by item appears in Appendix A. The instrument which was comprised mostly of closed-form questions with some open-ended ones, was divided into three sections. Part A, focused on characteristics of and

---

A notable proportion of CES is comprised of Francophone members. Faced with insufficient funds to translate questionnaires into French and a lack of information in the data base provided to the researchers concerning language of preference, the decision was made to forward the questionnaire to all randomly selected candidates. While many Francophone evaluators, particularly those working with the Canadian federal government, are bilingual, clearly the procedures adopted meant that several non English-speaking evaluators received the request to participate in the English-only survey. This regrettable decision should be understood as an attempt to maximize Francophone participation in the survey under prohibitive resource limitations. The authors deeply regret any offence that may have been taken.

It should be noted that from the original sample of 2,000 names, several were returned due to inaccurate postal addresses. Also, while precautions were taken to minimize the probability that members of more than one association would be sampled more than once, some twenty respondents received more than one questionnaire. In both instances, replacement evaluators were sampled from the appropriate association to maintain the target sample at 2,000.

Table 2.1: Population and Sample Characteristics

Association	(A) Number Available	(B) Target Sample	(C) Percent (B/A*100)	(D) Achieved Sample	(E) Percent (D/B*100)	(F) Response Option Pt B+C	(G) Response Option Pt A+B+C	(H) Response Option All+Nom.	(I) No. Prac-titioners	(J) Percent (I/H*100)
American Evaluation Association (AEA)	2,411*	1,000	41.5	306	30.6	114	126	66	36	54.5
Canadian Evaluation Society (CES)	1,308*	780	59.6	200	25.6	83	79	38	22	57.9
American Educational Research Association (AERA)	2,500	170	6.8	35	20.6	12	17	6	5	83.3
Assoc'n of Educational Research Officers of Ontario (AERO)	95	50	52.6	23	46.1	7	10	6	5	83.3
TOTAL	**6,314**	**2,000**	**31.6**	**564**	**28.2**	**216**	**232**	**116**	**68**	**58.6**

* Available in data base supplied by association.

BEST COPY AVAILABLE

perceptions about a specific collaborative evaluation project completed in the past three years. Respondents were invited to skip to Part B if they had not participated on such an evaluation. Several items included in Part A were drawn from questionnaires used in prior studies (Cousins & Simon, 1996; Cousins & Walker, 1995). Part B addressed a variety of views and opinions about evaluation in general and collaborative evaluation in particular. These were constructed based on evaluation theory and research. One set of items focussed on general issues about evaluation, while the remaining majority targeted items likely to be more directly relevant to collaborative evaluation. As discussed in chapter 1, most forms of collaborative evaluation are suited to formative evaluation problems, where the need for external, objective appraisals of program merit and worth, gives way to knowledge for program improvement and refinement. Scriven (1991) observes that this commonly used, improvement-oriented connotation of formative evaluation, departs from the intentions of the original definition of the term preparing for summative, decision-oriented evaluation. In the present case, it was assumed that respondents would be cuing on the improvement-oriented connotation of the term formative. While some of these items were drawn from prior work (Cousins & Walker, 1995), most were formulated specifically for the present survey.

Finally, Part C inquired into a minimal amount of relevant background information from respondents. Specifically, questions about the evaluator's organizational location, gender, training and experience were presented.

2.4 PROCEDURES

Permission to utilize as a basis for sampling membership mailing lists were obtained from each of AEA, CES, AERA, and AERO. Respondents were sampled and mailing labels produced. Each respondent received a survey package containing (1) the questionnaire booklet (colour coded for the respective association), (2) a covering letter addressed to evaluators (Appendix B1), (3) a sample covering letter addressed to practitioner participants (Appendix B2), and (4) a self-addressed stamped return envelope. Depending on their interests, evaluators were provided with three response options. First, the "anonymous" option required them to return the completed survey without revealing their identity. Respondents selecting this option completed the entire survey. A second group of respondents also answered anonymously but chose to omit Part A of the questionnaire. By implication this selection mean that they had not carried out a collaborative evaluation in the past few years. For both the anonymous groups requests for copies of a summary report of the findings could be made under separate cover or by other means (i.e., phone, fax, e-mail). A third response group selected the "confidential" option which asked respondents to nominate a representative program practitioner (non-researcher) with whom they collaborated on the evaluation identified in Part A. Respondents selecting the confidential option were to complete the form attached to the covering letter (see Appendix B1) and return it with their completed questionnaire. The information requested on the form was the name and complete mailing address of the program practitioner, the name of the evaluator, and an abbreviated name of the evaluation project on which both parties had participated.

Upon receiving this information, an identifying code number was assigned and program practitioners were then forwarded a survey package containing (1) a covering letter clearly identifying the name of the evaluator who had nominated them for the survey and an abbreviated name of the focal evaluation project, (2) a parallel version of the questionnaire adapted to program practitioners and (3) a self addressed, stamped return envelope. The code number was written on the practitioner questionnaire to ensure that the returned response could be linked to the responses provided by the evaluator. The program practitioner cover letter promised confidential treatment of the responses and a summary report when available.

Evaluator and program practitioner samples were each sent a follow-up letter three weeks after the initial mailing. Evaluators exercising the confidential option from whom responses had already been received were not sent the follow-up letter. Also eliminated from the reminder list were evaluators who had identified themselves as not being eligible for the survey by virtue of their lack of experience with or knowledge about collaborative forms of evaluation. Such individuals identified themselves through letters, telephone calls, or fax or e-mail transmissions. A master list was maintained in order to keep track of the status of evaluators sampled.[3]

2.5 DESCRIPTION OF EVALUATOR RESPONSES

Appendix A (page 8) shows the descriptive characteristics of the achieved sample of evaluators. In general, half were female with 12.5 years of experience as an evaluator on average. About half of the respondents claimed that evaluation was their primary responsibility and, on average, they claimed that 70% of the evaluations on which they have worked have involved program practitioner participants. Two thirds of the responding evaluators were located in public sector organizations of which 40% were universities. Less than 25% of the respondents were located in private sector organizations including consulting firms, non-consulting firms and self-employed operations.

Figures 2.1 through 2.6 show the breakdown of respondents by the response option selected. For these analyses, years of evaluation experience (Figure 2.3) and percentage of past evaluations considered to involve program participants (Figure 2.5) were categorized into high, moderate and low thirds to enable a chi-square test of statistical independence.[4] Figure 2.1 shows that evaluators located in universities and private sector firms were more likely to choose the confidential option whereas those located in public sector (non-university) settings were less likely than expected to do so, X^2 (10) = 19.21, p. < .05. Figure 2.2 shows that evaluators with fewer years of experience were more likely to report on a collaborative evaluation and to

---

This master was later destroyed once all the data had been entered on the computer and cleaned.

Cut points for years of experience were 7 years or less for low and 16 years or more for high. For percentage of collaborative evaluations the cut points were less than 50% for low, and greater than 95% for the high group.

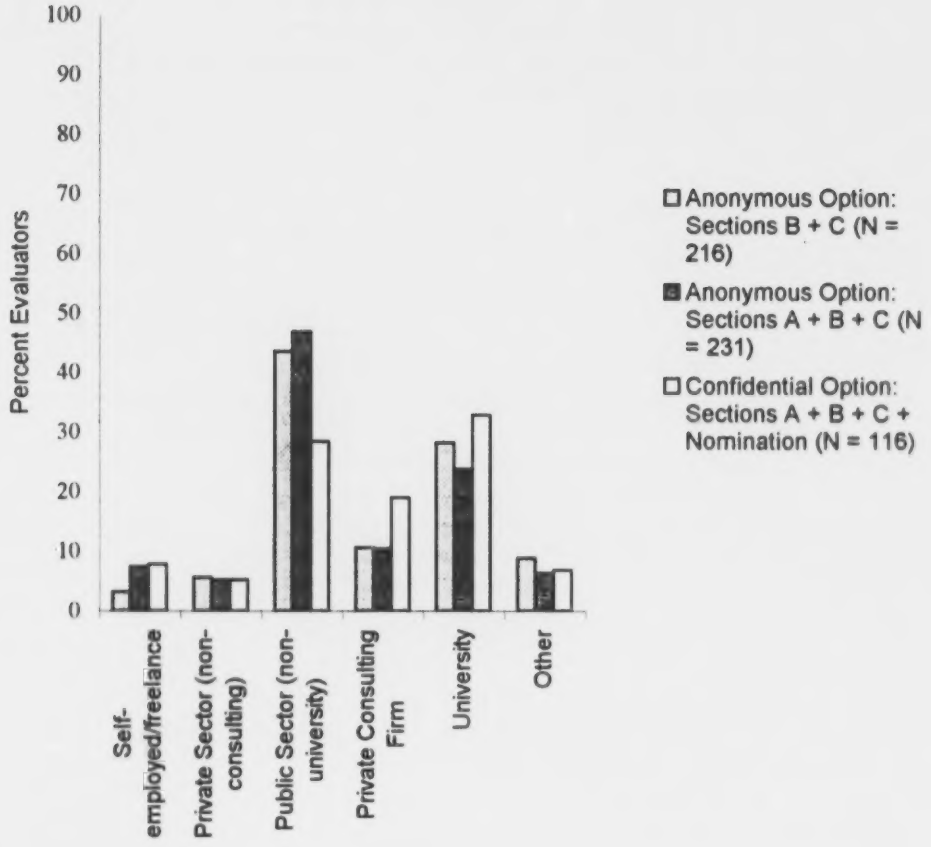


Figure 2.1: Evaluator organizational location by response option selected.

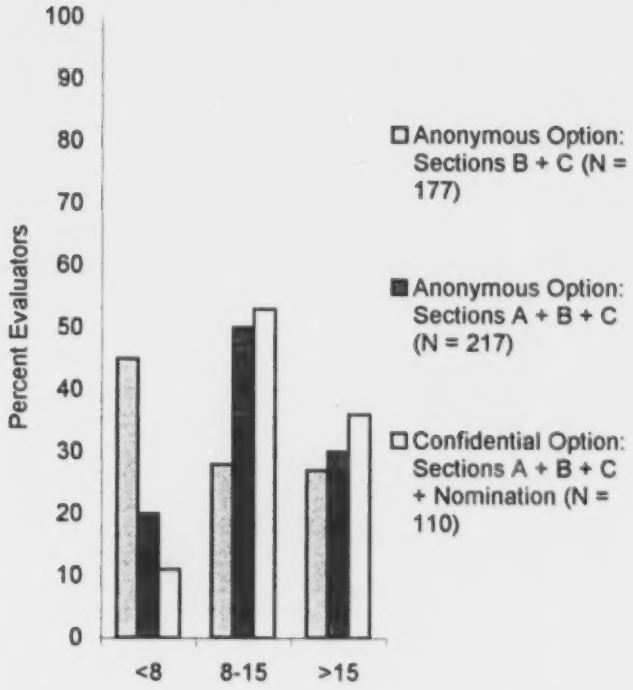


Figure 2.2: Evaluator years of experience by response option selected.

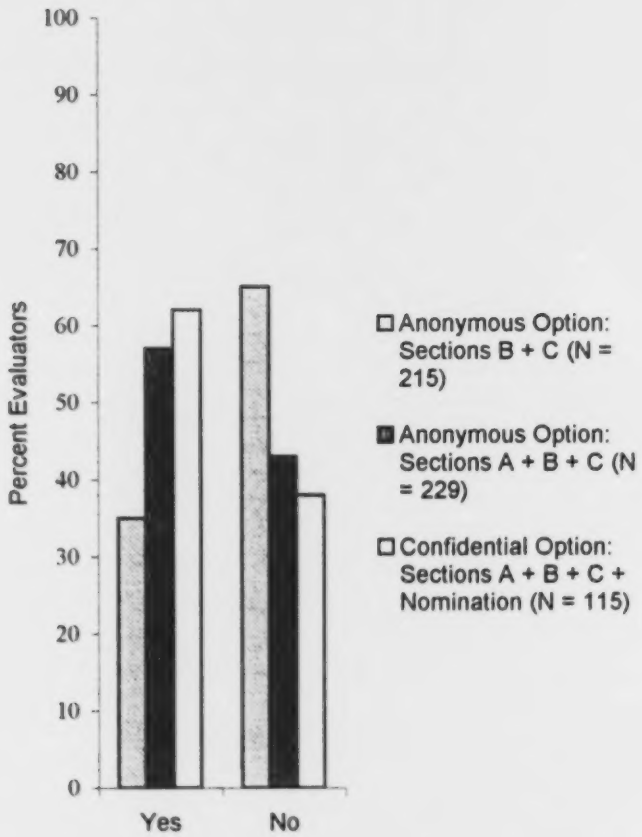


Figure 2.3: Claim to evaluation as primary responsibility by response option selected.

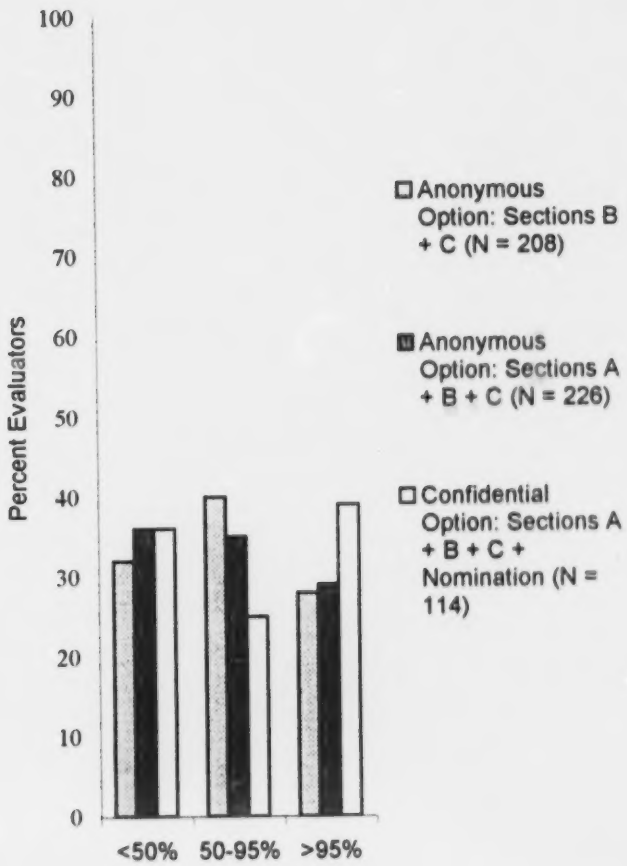


Figure 2.4: Percent evaluations considered collaborative by response option selected.

nominate a practitioner respondent than their more seasoned peers, X^2 (4) = 9.87, p. < .05. In Figure 2.3 we see that those who did not consider evaluation to be their primary responsibility were less likely to report on a recently completed collaborative evaluation X^2 (2) = 9.87, p. < .001. Finally, respondents who reported a higher proportion of the evaluations they conduct to be ones that involve program practitioners were more likely to reflect on one and to nominate a practitioner respondent, X^2 (4) = 23.39, p. < .001. There were no differences in chosen response option attributable to evaluator gender or training (highest degree achieved).

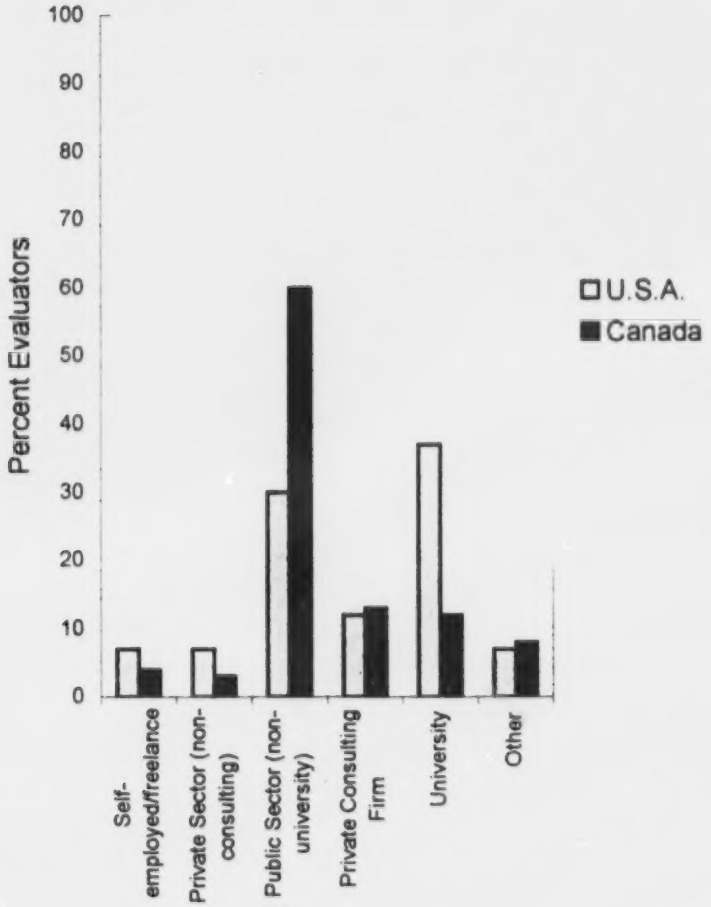


Figure 2.5: Evaluator organizational location by nationality.

A similar analyses was carried out to determine if primary membership in an American (AEA, AERA) versus Canadian evaluation association (CES, AERO)--a proxy for "nationality"--

would predict evaluator background characteristics.[5] The discretized versions of years of evaluation experience and percentage of past evaluations being collaborative were used. Figure 2.5 shows that two thirds of the American respondents were located in universities or other public sector organizations with slightly more being at universities. On the contrary, while slightly over two thirds of the Canadians were from these types of organizations, the vast majority of public sector evaluators were from non-university settings (likely governmental and para-governmental agencies), X^2 (10) = 64.42, p. < .001. Other categories were fairly similar for Canadians and Americans. No differences were found between Canadian and American evaluators in terms of their gender, primary responsibility or prior involvement in collaborative evaluation. Differences were observed, however, for years of experience as an evaluator, X^2 (2) = 22.05, p. < .001, and training, X^2 (2) = 22.05, p. < .001. American evaluators responding to the survey tended to be more well seasoned and more highly educated than their Canadian counterparts.

2.6 PLAN OF ANALYSIS

The research questions specified above and the format of the data were the primary criteria in selecting methods for analysis. All statistical analyses were conducted using SPSSpc software. Questions pertaining to descriptions of self-reported behaviors and opinions were answered by using descriptive statistical summaries. Much of this information appears in Appendix A. In order to enrich the description and to work toward reducing the large set of variables into a more readily interpretable form, several summary and scale variables were computed using a variety of techniques. In some cases exploratory factor analysis using varimax rotation was employed to identify patterns in the data. An eigenvalue of 1 was used as a criterion for factor selection and factors loadings exceeding .30 were considered meaningful.

Several scale variables were computed by taking linear combinations averages of sets of Likert type items. Item sets were determined on the basis of the conceptual framework for the study, or in the case of factor analyses, on the bases of factor loadings. Linear combinations were used instead of factor scores in order to maximize interpretability (i.e., average scale scores are on the same metric as component items; 1-4 for agree-disagree, 1-5 for frequency). Specifications for variable construction are provided in the ensuing sections of the monograph.

Relationships among variables were examined using a variety of bivariate and multivariate procedures. Intercorrelation matrices using Pearson correlation coefficients were examined for patterns among variable sets. Stepwise multiple regression was used to account for variation in criterion variables and to assess the relative uniqueness and magnitude of contributions by predictors. Repeated measures MANOVA was employed to test for differences between

---

[5] Clearly some members of American associations were Canadians and, to a lesser extent, vice versa, although the extent to which this was reflected in the achieved sample was minimal.

evaluator and practitioner groups. Between groups MANOVA was used to examine differences in responses attributable to demographic and background characteristics.

Finally, written comment data were typed and categorized by question. Items were content analysed and emergent patterns recorded.

CHAPTER 3: EVALUATOR VIEWS AND OPINIONS ABOUT COLLABORATIVE EVALUATION

A total of 564 evaluators provided their views and opinions about issues relevant to collaborative evaluation by completing Part B of the questionnaire. The reader will recall that the research questions of interest for this sample are:

R1. What are evaluators' opinions and views about collaborative evaluation? How variable are these impressions?

R2. Do demographic / background variables differentiate evaluators' views and opinions about collaborative evaluation?

These questions are addressed by first examining descriptive response patterns of evaluators at the level of individual items and, subsequently, after having reduced the items using factor analytic and scale variable construction techniques. Response patterns by item appear in Appendix A. Finally, relationships among scale variables and between these variables and demographic / background variables are considered.

3.1 DESCRIPTION OF RESPONSE PATTERNS

As shown in Appendix A, questions in Part B focused on general perspectives about evaluation, followed by a list of items associated with *formative* evaluation since the literature suggests that collaborative evaluation is best suited to improvement-oriented problems (Ayers, 1987; Cousins & Earl, 1992). All opinions were captured using a four-point agree-disagree Lickert-type scale. It should be noted that some evaluators expressed concern about responding to many of these items, the primary concerns having to do with vagueness of terms, the absence of a definition of collaborative evaluation and difficulties in rating in the absence of context. The frequency with which the 'not applicable' (N/A) option was selected is one indicator of these patterns. These caveats should be kept in mind as the findings are considered.

There were ten items under general perspectives on evaluation (see Table 3.1) with responses, on average, leaning toward the agreement end of the scale (i.e., exceeding 2.5). Variability was reasonably stable over items with standard deviations ranging from .71 to .82. The three items with which evaluators agreed the most were:

1.9 Evaluators should formulate recommendations from the study.

1.1 The evaluator's primary function is to maximize intended uses by intended users of evaluation data.

Table 3.1: Opinions about General Issues in Evaluation

Item	N	Response Distribution[a] SD	D	A	SA	Mean	SD	Rank[b]	95% C.I. Lower	Upper
Evaluator's primary function is to maximize intended use by intended users of the evaluation data.	539	3%	26%	42%	29%	2.98	0.81	2	2.91	3.05
More and more organizations are establishing internal evaluation capabilities.	514	2%	29%	54%	15%	2.82	0.71	4	2.76	2.88
Summative evaluations must be conducted by people external to the organization.	552	13%	54%	23%	10%	2.30	0.82	9	2.23	2.37
Formative evaluations are best done by internal members of the organization.	543	8%	48%	37%	7%	2.44	0.74	7	2.37	2.50
The evaluator's primary function is to maximize the technical quality of the evaluation.	546	4%	36%	44%	16%	2.72	0.78	5	2.65	2.79
Evaluators should be significantly involved in evaluation follow-up activities.	541	2%	22%	55%	21%	2.96	0.71	3	2.89	3.02
The evaluator's primary function is to maximize opportunities to bring about social justice.	505	20%	53%	22%	5%	2.13	0.79	10	2.06	2.20
Formative and summative evaluations cannot be separated.	527	8%	52%	31%	9%	2.41	0.76	8	2.34	2.47
Evaluators should formulate recommendations from the study.	544	1%	12%	54%	33%	3.19	0.67	1	3.13	3.24
Evaluators should have substantial expertise specific to the program being evaluated.	553	6%	49%	36%	9%	2.49	0.75	6	2.43	2.55

[a]SD = strongly disagree (1); D = Disagree (2); A = agree (3); SA = strongly agree (4)
[b]1 = highest agreement; 10 = lowest agreement

BEST COPY AVAILABLE

1.6 Evaluators should be significantly involved in evaluation follow up activities.

Two items with which evaluators tended to disagree were:

1.7 The evaluator's primary function is to maximize opportunities to bring about social justice.

1.3 Summative evaluations must be conducted by people external to the organizations.

These data suggest that evaluators, on average, believe that control of the evaluation should be maintained by them and that a commitment to enhancing utilization and responsiveness of evaluation ought to be adhered to (Patton, 1988). There also appears to be support for the notion that evaluation is moving away from external contracting toward reliance on internal research capabilities (Love, 1991). A commitment to fostering social justice as a primary function of evaluation was not evident in the sample.

Views about practitioner participation in formative evaluation showed considerable variation, with standard deviations ranging from .50 to .79 on the four-point scale. Average responses to these items appear in Table 3.2. Of the 17 items included in this section, the five with which evaluators agreed the most were:

2.15 Evaluation can help practitioners improve practice.

2.16 Evaluation can help practitioners to question fundamental beliefs and assumptions about practice.

2.17 Evaluation can result in fundamental changes in practice.

2.14 Evaluators should educate practitioners about the power and value of evaluation as a planned change strategy.

2.10 Practitioners participation in evaluation enhances the utilization of evaluation data.

There were no items in the list that evaluators tended to disagree with, on average, but on four items evaluators were divided in their opinions.

2.9 Practitioner's participation in evaluation enhances the technical quality of evaluations.

2.4 Special interest groups should participate in carrying out evaluation.

2.11 Practitioners' participation in evaluation helps to bring about social justice.

2.12 Evaluators should share control of evaluation projects equally with practitioners.

Table 3.2: Opinions about Practitioners' Participation in Formative Evaluation

Item	N	Response Distribution[a]				Mean	SD	Rank[b]	95% C.I.	
		SD	D	A	SA				Lower	Upper
The more stakeholder groups involved in evaluation the better.	547	3%	22%	53%	22%	2.94	0.74	10	2.88	3.00
Involving multiple stakeholder groups helps to offset political agendas.	531	2%	21%	61%	16%	2.91	0.67	12	2.85	2.97
Program beneficiaries should participate in carrying out evaluations.	542	3%	20%	58%	19%	2.92	0.72	11	2.84	2.98
Special interest groups should participate in carrying out evaluations.	509	7%	37%	50%	6%	2.54	0.72	14.5	2.46	2.60
People with a vital interest in programs should participate in carrying out evaluations.	542	3%	16%	61%	20%	2.98	0.70	9	2.92	3.04
People responsible for implementing programs should participate in carrying out evaluations.	544	1%	11%	62%	26%	3.12	0.63	7	3.05	3.19
Practitioners' participation in evaluation makes evaluations more efficient.	527	5%	36%	44%	15%	2.69	0.79	13	2.64	2.74
Practitioners' participation in evaluation makes research more responsive to local needs.	536	1%	9%	64%	26%	3.16	0.60	6	3.10	3.22
Practitioners' participation in evaluation enhances the technical quality of evaluations.	522	8%	47%	38%	7%	2.44	0.74	17	2.39	2.49
Practitioners' participation in evaluation enhances the utilization of evaluation data.	534	0%	5%	53%	42%	3.36	0.59	5	3.31	3.41
Practitioners' participation in evaluation helps to bring about social justice.	398	8%	45%	39%	8%	2.47	0.76	16	2.39	2.54
Evaluators should help train practitioners to do evaluations.	544	2%	14%	58%	26%	3.09	0.68	8	3.03	3.15
Evaluators should share control of evaluation projects equally with practitioners.	522	7%	43%	39%	11%	2.54	0.78	14.5	2.47	2.60
Evaluators should educate practitioners about the power of and value of evaluation as a planned change strategy.	544	0%	3%	55%	42%	3.38	0.57	4	3.33	3.43
Evaluation can help practitioners improve practice.	550	0%	0%	43%	57%	3.57	0.50	1	3.53	3.61
Evaluation can help stimulate practitioners to question fundamental beliefs and assumptions about practice.	549	0%	1%	48%	51%	3.50	0.53	2	3.46	3.54
Evaluation can result in fundamental changes in practice	546	0%	2%	53%	45%	3.44	0.53	3	3.39	3.48

[a]SD = strongly disagree (1); D = Disagree (2); A = agree (3); SA = strongly agree (4)
[b]1 = highest agreement; 17 = lowest agreement

BEST COPY AVAILABLE

These results highlight evaluators' views that collaborative evaluation is likely to have pragmatic consequences defined by influences on practice and especially on stimulating among practitioners reflection about fundamental dimensions of their practice. There was a strong sense that collaborative evaluation can make a difference in this regard. Influences on the technical quality of evaluation and on the political consequences of ameliorating social justice and attaining representativeness in collaboration through, for example, involving special interest groups appeared to receive less agreement from evaluators and more variability in the extent to which they agreed was observed.

As a final component in the description of evaluators' views and opinions about collaborative evaluation, written comment data in response to the item "Additional comments about evaluation practice" were content analysed (N=158). The emergent pattern revealed four categories of comments: (1) general views on evaluation practice (51.3%), (2) criticisms or concerns about the questionnaire (26.6%), (3) evaluator/practitioner views and (4) views about the collaborative evaluation reflected on in Part A. Representative comments under each category are presented below:

(1) General Views about Evaluation

> Evaluations are only effective when "the powers that be" are fully committed to the evaluation process and willing to implement manageable recommendations.
>
> "Timeliness" of evaluation is becoming a major issue in terms of the acceptance, impact and usability of evaluations. Getting evaluations shortened in time lines is essential.
>
> Motivation of the users is critical to evaluation's success and resulting change. Commitment and support of supervisory personnel is also essential
>
> Although only summative evaluation which assesses value fits Scriven's definition of "evaluation." I see formative evaluation as a powerful tool for program development and improvement; it encourages program staff to work collaboratively and gives all the opportunity to change direction if needed (even though this kills a strong research design!) "Social justice" is a concept that has no place in evaluation process. "Free from bias" is the way to go.

(2) Criticisms or concerns about the questionnaire

> I think it is critical to define one's concept of evaluation. It is a very diverse field. My sense is that your view of what evaluation is and entails is quite different from mine; and may lead to force fitting of responses.
>
> I do not know the difference between summative and formative evaluation.

You seem to assume that "social justice" is the goal of program evaluation. It is not! Evaluators may have the responsibility to provide information in order to make decisions about social justice. That may seem a nuance, but it is a fundamental one. PS: "What is social justice?"

Depositions of "summative" and "formative" should have been provided with the "B" section.

(3) Evaluator / Practitioner views

I have considerable experience involving practitioners in evaluation studies; it has never been a productive involvement. Practitioners seem to be incapable of providing impartial, independent advice to the evaluation team.

I think it is important that evaluators meet practitioners where they are and work with them to improve their ideas about evaluation and its uses and benefits. This does not mean that an evaluator can share responsibility "equally." It depends on their readiness and analysis.

There needs to be a strong and respectful relationship between practitioners and evaluators before good effects can come about. Evaluators cannot be distant researchers "reporting back" to practitioners. There must be good two-way communication.

(4) Views about the collaborative evaluation reflected on in Part A

I have worked on other projects in which participants were involved in the planning stages (i.e., method, instrument development). These were much more successful in terms of the usefulness of the data we collected and how the results were utilized -- they believed it was more useful because they helped to develop the project.

A strong project director is necessary to keep a project on course and not mixed in politics. Everyone--even the external evaluator--seemed to have an agenda.

Most recent evaluation project I did had such a meagre budget that there was no funding for involving practitioners. I worked on a contract basis. Evaluation would have been improved with more participation from schools.

Collectively, these comments reveal considerable variability in views and perspectives. Many comments about program practitioner involvement were quite favourable but problematic ones seemed to raise concerns about bias. This observation underscores the desirability to limit collaborative evaluation to improvement-oriented formative evaluation. Evaluators seemed somewhat perplexed about the connectedness of social justice to evaluation practice, a view that

has been shared before (e.g., Patton, 1994). Technical comments about the questionnaire resonate with the caveat identified previously.

3.2 SCALE VARIABLE CONSTRUCTION AND DESCRIPTION

In order to reduce the data items into meaningful constructs, factor analyses were carried out on each of the foregoing sets of items (i.e., question 1, 10 items; question 2, 17 items). The results showed that there existed too much error in variation among the general perspective items to warrant further analyses, and so these variables were dropped.

A more favourable technical outcome was observed for the views on practitioner participation in formative evaluation. The factor analysis yielded 5 factors accounting for 63.7% of the variability. Linear combinations of items for each of these factors produced acceptable levels of reliability (Cronbach's alpha) and so scale variables were constructed. Descriptive information about these five variables appears in Table 3.3. It may be observed that alpha for these variables ranged from .63 to .80 and that the percentage of variability accounted for in the factor analysis by each ranged from 28.3% to 6.3%.

Table 3.3 shows the component items for each scale variable. The factors, and therefore the scale variables, were interpreted as evaluators' opinions about (1) professional development for practitioners through participation, (2) empowerment of practitioners as a consequence of participation, (3) expected impact of collaborative evaluation, (4) practitioner participation in evaluation and (5) the maintenance of technical quality in collaborative evaluation. Evaluators tended to agree most readily with the view of collaborative evaluation as a professional development experience for practitioners and with the expected impact that such an approach is likely to have. They were relatively divided about the connection between collaborative evaluation and maintaining technical quality. Indeed the standard deviation for this latter scale variable was the highest of the five, showing that evaluators were not particularly consolidated in their views about this connection.

3.3 RELATIONSHIPS AMONG SCALE VARIABLES AND DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS

Table 3.4 shows the intercorrelations among the various categories of views and opinions. All correlations were positive and highly significant, ranging in magnitude from .25 to .84. The moderate size of most coefficients suggests that the scale variables are measuring different constructs. The largest correlation, .84, was at least partly attributable to overlapping component items in the participation and empowerment scales.

A multivariate analysis of variance was carried out in order to test for differences due to each of eight demographic / background characteristics. These variables, described more fully in

Table 3.3: Scale Variable Construction--Views and Opinions

Variable	Description	Alpha	Mean	SD	N
Opinion[a]					
Professional Development (12.5%)	Composite (average) of 5 items associated with professional development function of collaborative evaluation: (a) Evaluation can help stimulate practitioners to question fundamental beliefs and assumptions about practice; (b) Evaluation can help practitioners improve practice; (c) Evaluation can result in fundamental changes in practice; (d) Evaluators should educate practitioners about the power and value of evaluation as a planned change strategy; (e) Evaluators should help train practitioners to do evaluation. Scale is 1 (strongly disagree) to 4 (strongly agree).	.80	3.40	0.45	551
Empowerment (6.3%)	Composite (average) of 5 items associated empowering function of formative evaluation: (a) The more stakeholder groups involved in evaluation the better; (b) Evaluators should help train practitioners to do evaluation; (c) Evaluators should educate practitioners about the power and value of evaluation as a planned change strategy; (d) Special interest groups should participate in carrying out evaluation; (e) Program beneficiaries should participate in carrying out evaluation. Scale is 1 (strongly disagree) to 4 (strongly agree).	.74	2.99	0.51	550
Impact (9.9%)	Composite (average) of 3 items associated with consequences of practitioner participation: (a) Practitioners' participation in evaluation enhances the utilization of evaluation data; (b) Practitioners' participation in evaluation makes the research more responsive to local needs; (c) Practitioners' participation in evaluation helps to bring about social justice. Scale is 1 (strongly disagree) to 4 (strongly agree).	.63	3.05	0.64	398
Participation (28.3%)	Composite (average) of 7 items associated with practitioner participation in formative evaluation: (a) People with a vital interest in programs (e.g., program developers, sponsors, directors) should participate in carrying out evaluations; (b) People responsible for implementing or delivering programs should participate in carrying out evaluations; (c) Program beneficiaries should participate in carrying out evaluation; (d) Special interest groups should participate in carrying out evaluation; (e) Involving multiple stakeholder groups helps to offset political agendas; (f) Evaluators should share control of evaluation projects equally with practitioners; (g) The more stakeholder groups involved in evaluation the better. Scale is 1 (strongly disagree) to 4 (strongly agree).	.80	2.88	0.54	550
Technical Quality (6.3%)	Composite (average) of 3 items associated with evaluation quality control: (a) Practitioners' participation in evaluation enhances the technical quality of evaluations; (b) Practitioners' participation in evaluation makes evaluations more efficient; (c) Evaluators should share control of evaluation projects equally with practitioners. Scale is 1 (strongly disagree) to 4 (strongly agree).	.74	2.63	0.78	503

[a]Item selection for linear combinations was guided by an exploratory factor analysis if 17 opinion statements using varimax rotation. The percentage of variation explained by the factor appears in parentheses . Items were selected if the corresponding factor loadings exceeded .30. Items are rank ordered within category by loading.

BEST COPY AVAILABLE

chapter 2 are: response option (parts B + C; all parts; all parts plus practitioner nominated); nationality (American, Canadian); organizational location (5 possibilities); gender; years of experience as an evaluator (few, moderate, many); evaluation as prime responsibility (yes, no); percentage of evaluation projects involving practitioners (low, moderate, high); and training (doctorate, masters, other).

Table 3.4: Zero-order Intercorrelations among Opinion Variables--Evaluator sample (Pairwise deletion of missing data. N= 370 - 548)

Variable	1.	2.	3.	4.	5.
1. Professional Development	--				
2. Empowerment	.40*	--			
3. Impact	.25*	.65*	--		
4. Participation	.53*	.84*	.45*	--	
5. Technical Quality	.41*	.42*	.31*	.39*	--

p. < .001.

The results of these separate analyses revealed that only two variables--response option selected and percentage of evaluations involving practitioners--had differentiating effects. For response option Hotelling's multivariate F was 3.89 (df = 10, 752), p. < .001. Each of the univariate F's for the individual dependent variables was statistically significant at p < ,01 or less. A clear linear trend was evident: Respondents completing only part B and C were less in agreement with each of the opinion dimensions than those completing Part A as well (i.e., less supportive of collaborative evaluation). In addition, respondents who nominated a practitioner to be surveyed rated each of the dimensions the highest, on average. A similar trend was observed for the categorized version of percentage of past evaluations involving practitioners. Hotelling's F was 2.53 (df = 10, 638), p. < .01. Again, each of the five univariate F tests were statistically significant showing a clear trend of higher ratings being associated with greater prior involvement in collaborative evaluation. One demographic variable, nationality, yielded a statistically significant multivariate effect, Hotelling's F = 2.19 (df = 5, 378), p. < .05. However, the effect was weak and did not show up among the univariate F tests. Generally, Canadian evaluators appeared to be slightly more receptive to the dimensions of collaborative evaluation data represented by the scale variables. The opposite was true for the technical quality variable:

Americans tended to favour program practitioner involvement as a means of enhancing technical quality.

3.4 SUMMARY

Generally, evaluators' views and opinions about collaborative evaluation were favorable. Rating patterns revealed that respondents viewed practical consequences of evaluation--intended use for intended users, improvement of practice and enhanced understanding of evaluation as a planned change strategy--more favourably than arguments about improved technical quality of evaluation or collaborative evaluation's ability to bring about social justice. Comment data provided further support for this perspective, though some evaluators, likely those adhering to the view that the judgment of program merit and worth is evaluation's central purpose, raised concerns about the deleterious influences of practitioner biases and self-serving interests. Failure to provide in the questionnaire a sharp definition of formative evaluation as an improvement-oriented enterprise may have introduced some error into the responses.

Factor analyses showed that several underlying dimensions were evident in respondents' views. Average ratings of factor scale variables confirmed the pattern described above. Tests for the differentiating effects of demographic and evaluator background characteristics revealed that more favourable attitudes toward collaborative evaluation were to be found among those who engage in such activities more often and, to a slight degree, among Canadian evaluators, as opposed to their American counterparts.

CHAPTER 4: COLLABORATIVE EVALUATION PRACTICE

Almost two thirds of the sample (N = 348; 61.7%) provided descriptions and perceptions about a recently completed collaborative evaluation on which they had worked (see Appendix A, Part A). These data permitted answers to the following research questions.

R3. How do evaluators describe their collaborative evaluation projects? In what ways do they characterize its purposes and defining features? What types of impact and consequences do they report?

R4. To what extent are evaluators' reports about impact and consequences predicted by self-reported features of their projects? . . . their views and opinions about collaborative evaluation?

R5. Do demographic / background characteristics differentiate evaluators' descriptions and perceptions about the focal collaborative evaluation processes and consequences?

Again, the reporting strategy will be to describe numerical and written comment response patterns and then to construct a set of variables to be subsequently subjected to bivariate and multivariate tests for differences. The reader is referred to Appendix A for item level descriptive statistics.

4.1 DESCRIPTION OF RESPONSE PATTERNS

The reported studies lasted about one and one half years on average. Respondents were initially asked to describe the evaluation in terms of (a) purposes, (b) methodological features and (c) its main audience. The primary intention behind this question was to have respondents refresh their memories about the project and its consequences, however these data also provide some interesting insights into the nature and context of the collaborative evaluations. Written responses were provided by 335 respondents; these were content analysed within each of the three categories stated above.

Concerning purposes of the evaluation project, evidence surfaced to make the determination as to whether the project was improvement-oriented (formative), judgmental or decision-oriented (summative) or some mix of the two. The data were somewhat difficult to code and about 15% of the respondents did not provide sufficient information to make the determination about purpose. Of those remaining, elements of decision-oriented or summative aspects of evaluation were evident in almost two thirds of the responses (63.0%). (All evaluations specified as evaluations of program merit or worth were included in this category, including those that would fit the original conception of formative evaluation, i.e., those judged to

be mini-summative evaluations or in some sense preparatory for summative evaluation). Some examples were:

> Purpose: to determine affectiveness of new state program in reducing infant mortality.
>
> Purposes: Determine the merit and worth of features of a teacher education project.
>
> Assess the costs and performance of commercial lighting DSM programs.
>
> Purpose: to quantify replacement of barium enema by colonoscopy and assess these changes in relation to age and gender related risk factors, place of service, physician specialty and costs.

Several of the categorized responses (32.0%) were distinctly improvement-oriented. Examples of this category were:

> Participatory evaluation of an innovative educational program designed to facilitate middle school childrens' transition into high school.
>
> Formative evaluation of exhibit prototypes for a science museum major exhibit, studying instructional effectiveness with kids.
>
> Statewide evaluation of preschool programs for children with disabilities. Purposes: compliance monitoring and program improvement.
>
> Objective was to develop as national evaluation framework for ten provincial/federal service centres (single window access to multiple programs and services).

Finally, relatively few (5%) of the responses were mixed. That is to say, the information provided explicitly suggested the existence of both improvement-oriented and judgment-oriented purposes. For example:

> Two fold purpose: (1) document and describe implementation of collaborating effort between a school district and a state health/welfare dept., to locate a family assistance service on school site to provide services to students and families more effectively; and (2) assess outcomes of collaboration on service system and outcomes.
>
> Formative for improvement and establishing strategic direction.
>
> Formative and summative evaluation of an innovative preschool substance abuse prevention program.

The second part of this question inquired into methodological features of the evaluation. Written comments were loosely coded into exclusively qualitative, exclusively quantitative and mixed-method. A large proportion of the responses (37.0%) did not include information permitting this determination. The remainder were found to be relatively straight forward to code. Of them, 38.4% were exclusively quantitative designs, while 39.8% employed a mixed-method approach. Survey techniques were identified as the most frequently used quantitative approach, although some respondents described quasi-experimental designs. The remaining projects (21.8%) were judged to rely exclusively on qualitative methods, the majority of which were interview studies. Some such studies included document and archival data analysis. Very few reported using what might be termed participant observation approaches.

To differentiate the main audience of the evaluation project, responses were coded as program decision makers (including sponsors, managers, developers and implementors), program beneficiaries or a category combining the two (see Table 4.1). Again, over one third of the responses (36.4%) provided insufficient data to make this determination. Of the responses remaining, exactly two-thirds identified program decision makers and another 24.9% identified this group along with program beneficiaries. Few of the responses identified only program beneficiaries (9.4%).

Table 4.1: Frequency of Stakeholder Group Participation as Reported by Evaluators

Item	N	Response Distribution[a]					Mean	SD	Rank[b]	95% C.I.	
		Ne	R	S	F	A				Lower	Upper
Program developers.	305	13%	9%	22%	29%	27%	3.41	1.34	4	3.23	3.58
Program managers or directors.	327	5%	8%	22%	38%	27%	3.65	1.12	1.5	3.51	3.80
Program sponsors or funders.	285	30%	16%	25%	16%	13%	2.65	1.38	1.5	2.47	2.83
Staff responsible for implementation.	328	6%	6%	26%	33%	29%	3.64	1.20	3	3.48	3.80
Intended beneficiaries of the program.	320	29%	17%	27%	17%	10%	2.52	1.30	5	2.35	2.69
Special interest groups.	266	47%	16%	22%	9%	6%	2.09	1.23	6	1.93	2.25

[a]Ne = never (1); R = rarely (2); S = sometimes (3); F = frequently (4); A = always (5)
[b]1 = highest agreement; 6 = lowest agreement

Evaluators indicated that particularly large numbers of practitioners collaborated on the study; while the average was close to 20, this mean was inflated due to extreme cases--the median number of stakeholders involved in the project was 6. Table 4.2 (panel a) shows the sorts of activities in which stakeholders were participating. Generally, the pattern corresponds to the traditional stakeholder-based approach, with heavier levels of involvement in scope and design phases as well as activities centered around interpretation and dissemination of findings.

Table 4.2: Actual Participation in Evaluation Activities

A) Practitioner Participation

Item	N	Response Distribution[a]					Mean	SD	Rank[b]	95% C.I.	
		Ne	R	S	F	A				Lower	Upper
Defining the scope of the evaluation.	344	6%	3%	19%	31%	41%	4.01	1.10	1	3.89	4.13
Designing the study.	344	8%	12%	37%	23%	20%	3.35	1.17	3.5	3.22	3.48
Developing data collection instruments.	337	12%	15%	33%	20%	20%	3.19	1.26	7	3.06	3.33
Collecting data/information.	341	17%	11%	20%	25%	27%	3.33	1.43	5	3.17	3.48
Processing and preparing data for analysis.	338	45%	24%	14%	7%	10%	2.14	1.33	10	2.00	2.29
Analyzing data.	339	45%	20%	17%	9%	9%	2.17	1.34	9	2.02	2.31
Interpreting results.	339	16%	12%	25%	26%	21%	3.23	1.34	6	3.09	3.38
Preparing reports for dissemination.	37	28%	19%	24%	12%	17%	2.72	1.42	8	2.56	2.87
Formulating recommendations from the study.	331	14%	15%	22%	20%	29%	3.35	1.40	3.5	3.20	3.51
Disseminating results to and recommendations to intended users or audiences.	326	7%	11%	21%	30%	31%	3.68	1.23	2	3.54	3.81

B) Evaluator Participation

Item	N	Response Distribution[a]					Mean	SD	Rank[b]	95% C.I.	
		Ne	R	S	F	A				Lower	Upper
Chairing project meetings.	325	21%	8%	20%	28%	23%	3.26	1.44	9	3.10	3.42
Providing guidance about technical research matters.	339	1%	2%	10%	39%	48%	4.29	0.84	3	4.20	4.38
Developing data collection instruments.	340	2%	5%	12%	27%	54%	4.27	0.98	4	4.17	4.38
Collecting data.	339	11%	12%	14%	21%	42%	3.74	1.39	7	3.59	3.89
Processing and analyzing data.	340	2%	5%	9%	23%	61%	4.36	0.99	2	4.25	4.46
Preparing reports for dissemination.	338	1%	3%	7%	24%	65%	4.49	0.84	1	4.39	4.58
Formulating recommendations.	338	3%	4%	14%	27%	52%	4.20	1.03	5	4.09	4.31
Disseminating results to intended users or audiences.	334	7%	8%	21%	27%	37%	3.80	1.22	6	3.67	3.93
Helping practitioners to develop technical research skills.	321	14%	20%	27%	18%	21%	3.11	1.33	10	2.96	3.26
Educating practitioners about the power and value of evaluation as a planned change strategy.	324	9%	10%	23%	29%	29%	3.59	1.26	8	3.45	3.73

[a]Ne = never (1); R = rarely (2); S = sometimes (3); F = frequently (4); A = always (5)
[b]1 = highest agreement; 10 = lowest agreement

BEST COPY AVAILABLE

However, participation in instrument development and data collection activities was also reported to be common. Disseminating results and recommendations was the single most frequently identified activity for practitioners on average. Practitioners were rarely involved in the technical data processing and analysis activities according to evaluators.

Two thirds of the evaluators indicated that participating stakeholders belonged to more than one group. A quick perusal of comment data revealed considerable variation. Listed were program managers, senior organization managers, program staff, classroom teachers, principals, education consultants, state level government administrators, advocacy groups, program beneficiaries, and so on. Evaluators indicated that most of the participants belonged to stakeholder groups affiliated fairly directly with the program; developers, managers, funders or implementors (see Appendix A, Part A, Items 6.1-6.7). These categories correspond to Alkin's (1991) classification of primary users, those who have a vital interest in the program and who might be expected to be able to act on data once available. Chief among them were program managers and implementors. Intended beneficiaries were involved in the evaluations to a modest degree and special interest groups were rarely involved.

The issue of balance of control over evaluation project decision making was measured in two ways. First, evaluators were simply asked who was in control, researchers, practitioners or some balance of the two groups. This item was not particularly discriminating; almost two-thirds of the respondents indicated a balanced approach, with over a quarter leaning toward researchers and under 10% toward practitioners (see Figure 4.1). A second way to get at the issue of control was to ask evaluators about their own level of participation in a variety of research tasks. Table 4.2 (panel b) shows the pattern of involvement of evaluators according to their self-reports. A caveat with this measure is that it cannot necessarily be assumed that a respondents' non-involvement in a given task (e.g., project meeting chair) implies that practitioners assumed responsibility for that activity; another researcher on the evaluation team may have had the responsibility. Bearing this warning in mind, it may be seen in Table 4.2 that aspects of the role most frequently carried out by evaluators were providing guidance on technical research matters, developing data collection instruments, processing and analysing data, reporting and formulating recommendations. Each of these averaged out between "frequently" and "always" on the five-point frequency scale. Evaluators were least involved, on average, in training practitioners to do research and chairing project meetings. Persuading practitioners about the power and potential of evaluation was also an activity in which evaluators engaged regularly. It can be noted in Table 4 that there was considerable variation among evaluators; standard deviations on the five-point scale ranged between .84 and 1.44.

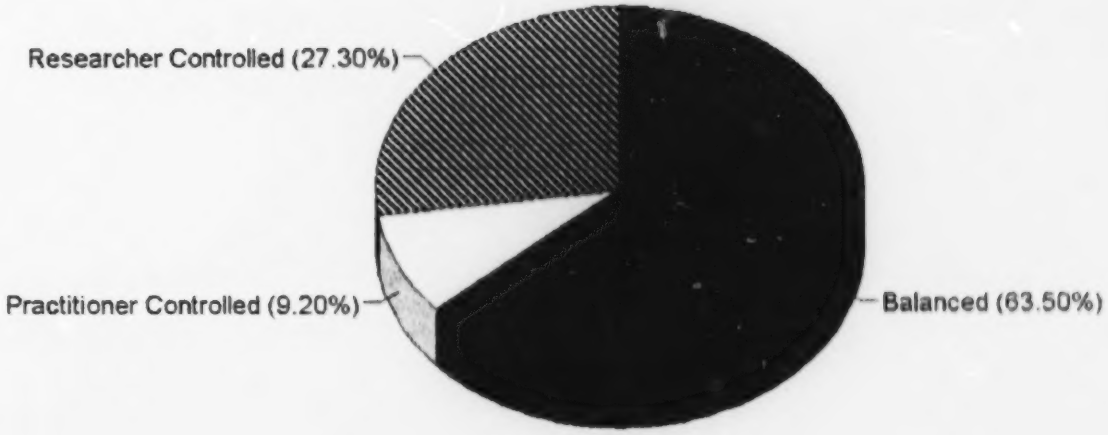


Figure 4.1: Control of evaluation project decision making as reported by evaluators (N = 337)

Evaluators were asked how practitioners came to be involved in the project and how they, themselves, joined the team. Of the 315 evaluators responding, 75 provided information which could not be coded. However, of the remaining 240, 27.1% indicated that practitioners were assigned to the project, 25.0% were recruited or selected by the researcher(s), 22.9% volunteered for participation, while the remaining 27.1% had mixed reasons. Sometimes different reasons pertained to different individuals within projects. Fourteen (4.3%) of evaluators provided data that were not readily codable in terms of reasons for their own involvement. Of the remaining 312 answering this item, the most frequent response was that their involvement was considered to be part of their job (65.4%). Some examples were:

> Work related. I am the county "evaluator." I am shared around as needed.
>
> Through a cooperative agreement with the state we provided technical assistance relating to program and evaluation services.
>
> As Director of Evaluation I oversee the conduct of most evaluation studies. I work very closely with program managers frequently.

Another way in which they became involved included their being recruited or hired (contractually) (29.8%) for the project. As the following comments show, often, the evaluator was known professionally to the sponsoring agency ahead of time:

> Project was awarded through an open competition. I held several informal meetings with the practitioner group before the evaluation started. It helped that I had worked with one of them a few years before this project.
>
> State-trained ("approved") evaluator for special innovation grants funded through Chapter II. I was recommended by state dept of education representatives.
>
> The program was housed in my university and the program developers knew of my work. They approached me; it was a good fit.

Finally, some evaluators volunteered to help out. They gave of their time for a variety of reasons, most frequently to gain experience or access to data for graduate studies, but in some cases a certain amount of choice and selection was operative.

> I carried out a practicum at the health unit in the summer of 1994 and continued with the unit for a thesis topic for my Master's requirements. The practicum involved the beginning stages of the evaluation.
>
> Became involved in planning the evaluation as part of Master's thesis. Remained involved to implement evaluation. Previous work experience with program management.
>
> I volunteered when I was between jobs. Most of us have been on the committee since its inception and have developed a substantial skill level in evaluating the wide range of Heart Health initiatives. We have all upgraded in the field of evaluation through university extension courses or short workshops.
>
> Volunteered or chose projects to work on early in 1995.

In addition to inquiring about the composition of the team and how the collaborative arrangement came about, evaluators were also asked to provide an indication of the general methodological epistemological approach taken. That is to say, was a conceptual framework employed, and if so was it specified in advance of data collection (pre-ordinate) or during or after data collection (emergent)? The results displayed in Figure 4.2 show that in over 90% of the cases a conceptual framework was employed, with the majority being developed early in the research process.

Turning to issues of dissemination and the communication of findings, evaluators were asked to indicate who received the recommendations and results of the study and how they were communicated. Tables 4.3 and 4.4 show a fair amount of diversity in responses. At least three quarters of the evaluators identified program developers, managers or sponsors and implementors as recipients. The reader will recall that individuals from these groups also usually participated in

carrying out the evaluation. Less frequently cited were intended beneficiaries, special interest groups and academic audiences, but these were identified by about 40% of the respondents. Table 4.4 shows that evaluation teams tended to use conventional written (executive summary, technical report) and oral presentation formats. Eighty percent or more of the evaluators claimed this to be the case. Less often implemented were follow-up committee meetings, and the more fluid written communications such as newsletters and circulars (24.6%).

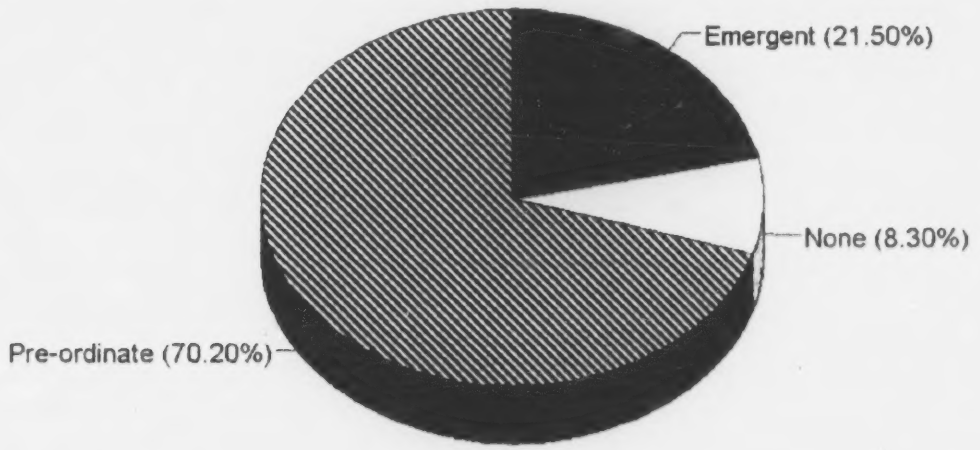


Figure 4.2: Specification of conceptual framework as reported by evaluators (N = 339)

Table 4.3: Audiences for the Evaluation as Reported by Evaluators

Item	N	% Reporting	Rank[a]
Program developers.	337	73	4
Program managers or directors.	337	91	1
Program sponsors or funders.	337	76	3
Implementors of the program.	337	82	2
Intended program beneficiaries.	337	43	5
Special interest groups.	337	40	7
Academic audiences.	337	41	6

[a]1 = most frequently reported, 7 = least frequently reported

Table 4.4: Modes of Communication as Reported by Evaluators

Item	N	% Reporting	Rank[a]
Executive summary.	337	82	2
Technical written report.	337	90	1
Newsletter/communication/circular.	337	25	5
Oral presentation(s).	337	80	3
Follow-up committee meeting(s).	337	48	4

[a]1 = most frequently reported; 5 = least frequently reported

Several items asked respondents to comment on their perceived impact of the evaluation project. Prior to direct, closed-form questions they were provided with an opportunity to summarize impact in their own words. Three-hundred and nineteen evaluators availed themselves of this opportunity. Of that group, 19.7% suggested that it was still to early to tell just what impact had occurred. Of the remaining responses, three distinct categories emerged. First, the majority of evaluators (56.6%) reported significant impact and identified specific changes that had come into place. Some examples were,

> I haven't really been involved since the study ended. The whole library structure has undergone radical changes since then. I still run into people who say they appreciated the work and have said "We're still talking about the issues you raised in your report." Concrete action? I don't know.

> The agency that administers the Medicare program is using the study's findings and recommendations to change the way it reimburses ambulance bills. Prior to the study, the agency did not know the extent to which it had been paying for uncovered ambulance transports.

> Major changes have occurred both in the program culture (i.e., approach to subject) and in actual policy and activities.

> The study has had substantial impact on local practice in schools. Changes in the state dept of education have meant limited attention to results by leadership. Changes in personnel at the Department of Education have also limited continuity and commitment to project by agency staff.

> Study resulted in modification of delivery (for staff) for better communication across delivery groups about the whole program, program changes for adults and children (beneficiaries). Results used to plan next year's program. Participation of all key groups and director enhanced impact.
>
> New curriculum materials were greatly influenced during the formative evaluation. New type of curriculum adoption workshop plan developed from summative eval.

Many respondents also provided insights about the study's impact and factors influencing it. While most comments about impact were limited to program related consequences, a number of evaluators alluded to organizational impact. A second category that emerged provided claims of impact but no real specification of concrete actions or effects of the evaluation. The following examples are illustrative.

> Results have been used to set annual program goals and objectives. They have helped to clarify program outcome measures. On a daily level, however, the results have had only minimal impact. Residential treatment has so many components that change is difficult.
>
> Proved to public health nurses that evaluation of awareness campaigns (traditionally viewed as impossible to evaluation) can and should be done. Results will be used to develop template for using this model of change in health promotion (currently being tested with another heart health project). Blood pressure awareness results will be used to prompt further work in this area. Short-term nature of project (1 year) limits ability to respond to recommendations.
>
> The study clearly affirmed the value of co-op ed and identified obstacles to use, especially for advanced students. The report is still praised but the program management and scope has been reduced since the Board experienced severe fiscal restrictions and reduced programs that were not mandated. Evaluation is only one consideration in difficult fiscal decisions.
>
> The program has been expanded. Findings were that program had been effective, but expansion should be slow and deliberate. Decision making in public schools doesn't work that way. Expansion probably too much, too soon.

A final category of comments showed that only limited impact had occurred. Examples were:

> Unfortunately, although the program management have been very supportive, apart from dissemination of results, there have not been the dedicated staff to follow-up with implementation of recommendations. It is unknown at this time, to what extent recommendations have indeed been implemented.
> Encouraged several people to work on doctoral degree. Established alternative school as

successful only if we treat at-risk students with respect and concern. A new superintendent of schools limited the impact of the study. The study was "not his baby." Could not identify with the project.

Not as much as we would have hoped, because changes in the health services delivery environment overtook the study. We learned a great deal about the cost of cuts in service delivery on STD prevention. As is so often the case when results of evaluations can't be acted upon for political reasons, COC has responded by initiating more research.

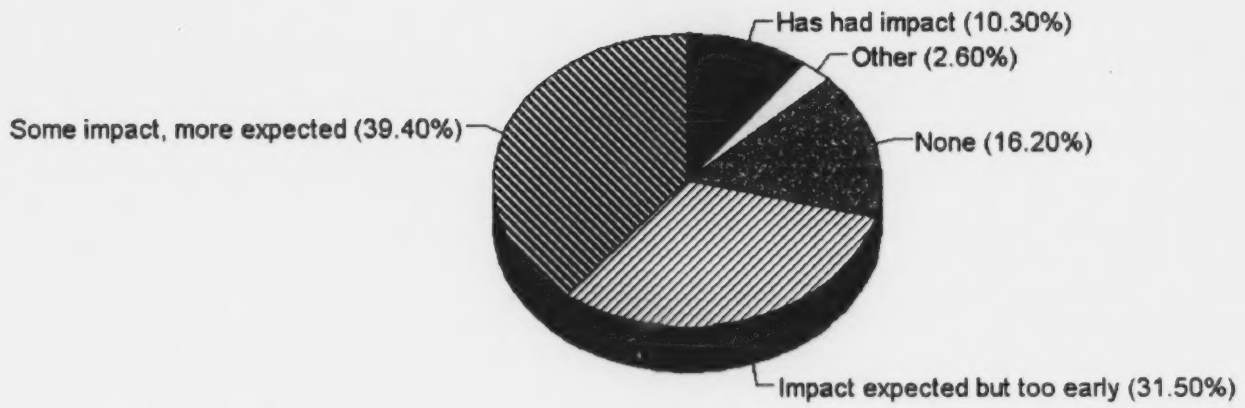
This study is just drawing to a close and the final report is being written. Preliminary & Process Reports have been used to: support continued funding of projects, design new initiatives by the funding agency, enlarge projects at the six sites, improve participating agency buy-in at the six sites (they were mandated to develop conditions), and to test new research methods. Study impact was limited by the inexperience of project site directors. Also limited by staff changes at the funding agency.

Study had limited impact given fiscal restraints imposed during and after the evaluation. The model is innovative, however, and appeals to practitioners. Without additional resources, it is unlikely all recommendations can be implemented.

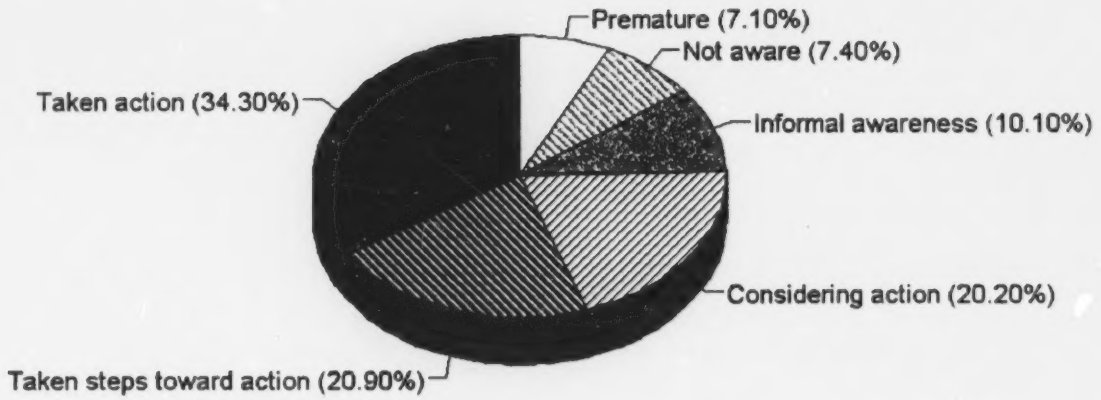
Study has permitted the organizers to better define their future activities. Impact was limited because of the low availability of the organizers to disseminate the results.

Evaluation impact and utilization was assessed in a variety of ways using closed form items. First, evaluators were asked to indicate the extent to which impact had been observed. A shown in Figure 4.3 (panel a) less than 17% revealed that no impact had been observed, while another about 30% suggested that impact was expected but it was too early to tell at the time of the present survey. Fifty percent of the evaluators perceived that some impact had occurred, and in many cases, more could be expected. Figure 4.3 (panel b), on the other hand, reports the finding of a scale of utilization based on prior work by Larson (1981). The scale ranges from "premature" to observable actions taken as a result of the evaluation. About one quarter of the respondents indicated that either it was too early to tell, intended users were not aware of findings or their level of awareness was based on informal information only. On the other hand, several believed that the intended users were currently considering the evaluation data (20.2%), or were either considering steps toward (20.9%) or taking action (34.3%) based on the data. Results displayed on this scale are consistent with those in panel a.

Table 4.5 shows a set of items that correspond to the type of impact that may have been observed as a consequence of the evaluation. With the exception of practitioners developing their skills in doing research (mean = 2.6) the remaining items reflected phenomena observed at least "sometimes" on average according to the frequency scores assigned. The conventional instrumental (discrete decision) and conceptual (learning) outcomes typical to most evaluation



a) Perceived impact (N = 340)



b) Intended use by intended users (N = 326)

Figure 4.3: Perceived impact and intended uses of evaluation data

Table 4.5: Perceived Consequences of the Evaluation

Item	N	Response Distribution[a]					Mean	SD	Rank[b]	95% C.I.	
		Ne	R	S	F	A				Lower	Upper
Intended users have based (or will base) significant decisions on this information.	314	2%	7%	28%	44%	19%	3.70	0.93	2	3.60	3.81
Intended users have learned (or will learn) about their practice.	308	1%	5%	31%	43%	20%	3.76	0.88	1	3.66	3.86
Intended users have developed (or will develop) their research skills.	293	13%	36%	34%	12%	5%	2.60	1.02	7	2.48	2.72
Intended users have learned (or will learn) that evaluation can be a powerful and valuable planned change strategy.	319	3%	12%	38%	33%	14%	3.43	0.98	4.5	3.32	3.54
Data have helped (or will help) intended users incrementally improve their performance.	300	3%	11%	36%	38%	12%	3.44	0.95	3	3.33	3.55
Data have helped (or will help) intended users to question basic assumptions and beliefs about their practice.	312	3%	9%	42%	33%	13%	3.43	0.94	4.5	3.32	3.54
Data have helped (or will help) stimulate fundamental changes in practice.	312	6%	15%	40%	28%	11%	3.24	1.03	6	3.13	3.36

[a]Ne = never (1); R = rarely (2); S = sometimes (3); F = frequently (4); A = always (5)
[b]1 = highest agreement; 7 = lowest agreement

utilization frameworks were both rated relatively high (approaching "frequently"). But several other changes and influences were also reported to have taken place. These indicators are associated with organizational learning phenomena or organization change patterns. Evaluation that was thought by evaluators to have been taken more seriously by intended users, have led to incremental improvements in performance, and to have generated fundamental questions about assumptions and practices. Intended users were perceived to have developed some sense of the extent to which evaluation could contribute within the organization.

Given its non-traditional status among approaches to evaluation, collaborative evaluation can be a trying experience even for seasoned evaluators. Figure 4.4 shows that, from an affective point of view, evaluators seem to take the process in stride, and generally, perceived it to be a positive experience. The least favourable rating was associated with the level of stress generated by the project, but all other indicators were positive. Evaluators found the experience to be particularly rewarding and favourable, though somewhat inefficient and frustrating.

4.2 VARIABLE CONSTRUCTION AND DESCRIPTION

As with the section on views and opinions (chapter 3) it was desirable to reduce the data set by constructing a set of variables for use in subsequent analyses. Variable construction and descriptive statistics are described in Table 4.6. These are grouped into dependent variables (measures of impact) and collaborative evaluation process variables.

There were six impact variables, the first two--expected use and level of use--taken as responses to direct closed-form questions on the survey instrument. The remaining impact variables were linear combinations (averages) of questionnaire items all judged to be satisfactorily reliable. The final impact variable corresponds to the evaluators' perceived impact of the project on his or her own affect. The descriptive statistics suggest that perceived impact was consistently favorable across the various indicators.

Process variables correspond to the three key dimensions identified in the conceptual framework. Each variable was a linear combination of ratings and judged to be reliable. Evaluator control was derived from variables associated with the evaluators' involvement in the research process. Stakeholder diversity is an aggregate of ratings of frequency of involvement by six different stakeholder groups ranging from program sponsors to implementors to beneficiaries. Finally, depth of participation corresponds to ratings of stakeholder participation in specific evaluation tasks. On average stakeholder diversity and depth of participation correspond to 'sometimes' on the five-point frequency scale while evaluator control was closer to 'frequently'.

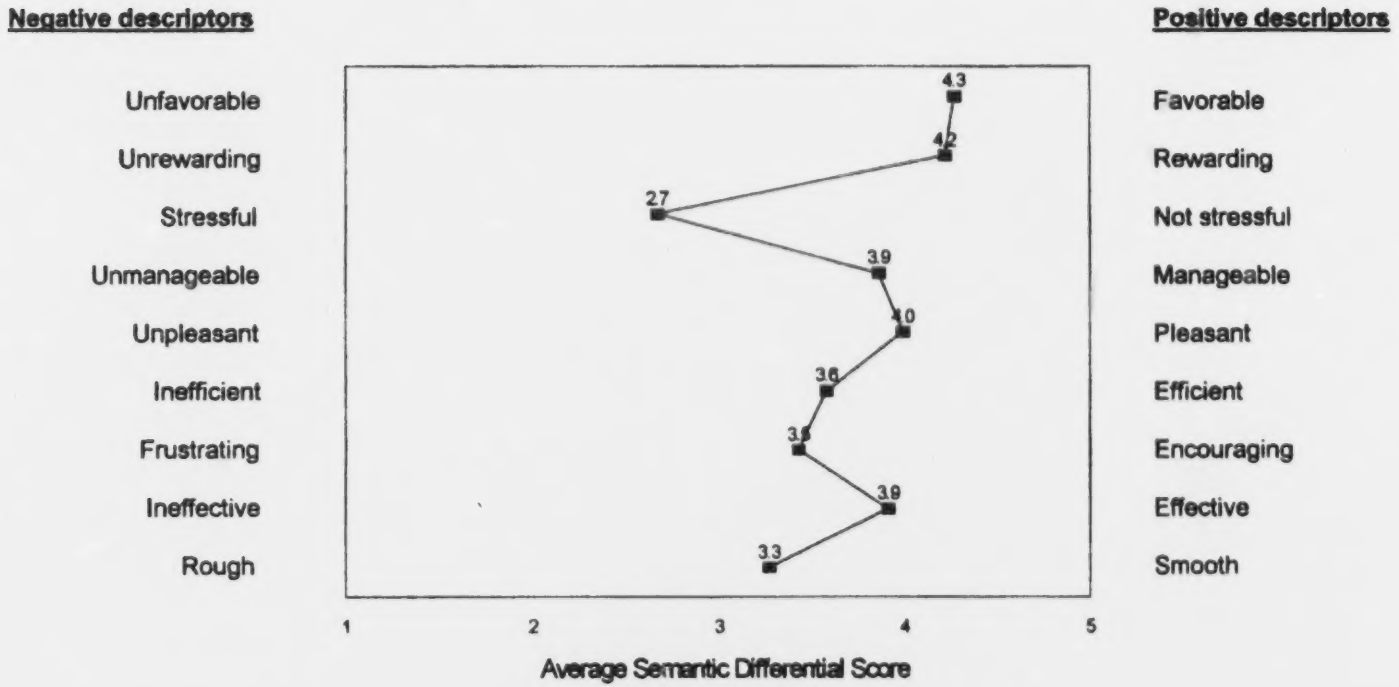


Figure 4.4: Perceived affective impact on evaluators (N = 335 - 340)

Table 4.6: Variable Construction and Descriptive Statistics

Variable	Description	Alpha	Mean	SD	N
Impact					
Expected Use	Expectations for utilization of evaluation data. Range is 1 (premature) to 6 (action being taken).	NA	4.44	1.56	326
Level of Impact	Four-point scale corresponding to levels of impact. Scale is 1 (no impact) to 4 (impact).	NA	3.39	0.96	340
Instrumental Use	Composite (average) of frequency scores for each of 3 observations of behavioral change: (a) Intended users have based (or will base) significant decisions on this information; (b) Data have helped (or will help) intended users incrementally improve their performance; (c) Data have helped (or will help) stimulate fundamental changes in practice. Scale is 1 (never) to 5 (always).	.78	3.46	0.81	283
Conceptual Use (Substance)	Composite (average) of frequency scores for each of 2 observations of program-related learning: (a) Intended users have learned (or will learn) about their practice; (b) Data have helped (or will help) intended users to question basic assumptions and beliefs about their practice. Scale is 1 (never) to 5 (always).	.60	3.59	0.77	299
Conceptual Use (Process)	Composite (average) of frequency scores for each of 2 observations of evaluation process learning: (a) Intended users have developed (or will develop) their research skills; (b) Intended users have learned (or will learn) that evaluation can be a powerful and valuable planned change strategy. Scale is 1 (never) to 5 (always).	.73	3.02	0.89	290
Affective Experience	Composite (average) of semantic differential scores concerning evaluator satisfaction with the collaborative experience. Scale is 1 (negative) to 5 (positive).	.88	3.69	0.75	338
Process					
Evaluator Control	Composite (average) of 10 frequency scores corresponding to evaluator participation in evaluation activities (e.g., chairing meetings, planning, instrument development, data analysis, reporting, disseminating). Scale is 1 (never) to 5 (always).	.79	3.92	0.63	338
Stakeholder Diversity	Composite (average) of 6 frequency scores corresponding to stakeholder groups participating in the evaluation (i.e., program developers, managers, sponsors, implementors, beneficiaries, special interest groups). Scale is 1 (never) to 5 (always).	.72	3.10	0.83	315
Depth of Participation (alpha = .88)	Composite (average) of 10 frequency scores corresponding to practitioner participation in evaluation tasks (e.g., planning, instrument development, data collection, data analysis, reporting, disseminating). Scale is 1 (never) to 5 (always).	.88	3.11	0.92	339

BEST COPY AVAILABLE

4.3 RELATIONSHIPS AMONG IMPACT, PROCESS and OPINION VARIABLES

Table 4.7 is an intercorrelation matrix showing associations among all pairwise comparisons of variables. At least six observations about this table are remarkable. First, while impact variables correlate reasonably high with one another the coefficients are of moderate size supporting the conclusion that the variables are measuring different aspects of impact. Second, level of impact does not correlate with most of the other impact variables and any of the process or opinion variables raising doubts about its sensitivity (reliability) as a measure of impact. Third, process and impact variables consistently intercorrelate with one another but the moderate size of coefficients allays concerns about multicollinearity in ensuing regression analyses; theses variables, too, are measuring different dimensions of process and opinion. Fourth, the variable measuring evaluator control is not associated with opinion variables. Fifth, process variables are more highly associated with those measuring impact than are opinion variables. Generally, expected use and level of use were unrelated to the predictors. Finally, all statistically significant zero-order correlations are positive.

The extent to which perceived impact depends on the unique influences of the self-reported process variables and opinions about collaborative evaluation was examined using stepwise multiple regression. Six models were run, corresponding to each of the impact variables. In each model, the eight process and opinion variables were specified as potential predictors. Table 4.8 reports the results of these regressions listing the predictors that explained a significant portion of variation in impact while controlling for the effects of other variables.

Though the predictors were unable to explain variation in the dubious level of impact variable, statistically significant models were fit in each of the remaining cases. The criterion variables most readily explained by the predictors were the measures of collaborative use, both that associated with learning about the program under evaluation (substance) and that relating to skill development (process). In each case about 30% of the variation was explained, most of which was attributable to the three key dimensions of process. Intended users appeared to learn more about their programs when a wider group of stakeholders participated, when evaluators were controlling the project and when stakeholder participation in technical activities was substantial. Similarly, intended users were more likely to develop research skills and an appreciation of evaluation as a strategic planning process when these three conditions were present. Also predictive of this sort of conceptual impact was the evaluators' favourable opinion as to the likelihood of impact associated with collaborative projects.

Diversity of stakeholder involvement in the project was related to two other sorts of perceived impact. First, evaluators reporting greater diversity were more likely to suggest that instrumental consequences of the evaluation were evident. This variable alone, explained 18% of the instrumental impact dependent variable. Five percent of the variation in evaluation satisfaction was also attributable to stakeholder diversity. As a greater range of stakeholders participated, the affective experience for evaluators appears to have been enhanced. Finally, a weak effect was

Table 4.7: Zero-order intercorrelations among Impact, Process and Opinion Variables (Pairwise deletion of missing data, N = 211 - 336)

	IMPACT						PROCESS			OPINION				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IMPACT														
1 Expected Use	—													
2 Level of Impact	.25***	—												
3 Instrumental Use	.28***	.04	—											
4 Conceptual Use (Substance)	.08	.06	.75***	—										
5 Conceptual Use (Process)	.20***	.08	.63***	.58***	—									
6 Affective Experience	.17***	.08	.48***	.38***	.43***	—								
PROCESS														
7 Evaluator Control	.11*	.07	.24***	.27***	.29***	.17***	—							
8 Stakeholder Diversity	.07	.01	.39***	.37***	.37***	.26***	.23***	—						
9 Depth of Participation	.05	.04	.26***	.27***	.38***	.20***	.12**	.43***	—					
OPINION														
10 Professional Development	-.06	-.08	.03	.10*	.17**	.05	.10*	.20***	.24***	—				
11 Empowerment	-.01	-.02	.18***	.12*	.25***	.04	.09	.37***	.34***	.40***	—			
12 Impact Expectation	-.06	-.01	.21***	.17***	.28***	.06	.04	.18***	.17***	.25***	.65***	—		
13 Participation	-.01	-.04	.17**	.13**	.17**	.14**	.06	.42***	.27***	.53***	.84***	.45***	—	
14 Technical Quality	-.08	.01	.14*	.10	.17**	.10*	.09	.36***	.14**	.41***	.42***	.31***	.39***	—

* $p < .05$, ** $p < .01$, *** $p < .001$

BEST COPY AVAILABLE

noted for evaluator control of the evaluation as a prediction of expected use. Only 2% of the variation in impact of this sort was explained.

Table 4.8: Stepwise Multiple Regression of Impact on Process and Opinion

Criterion	R^2	df	F	Variables remaining in the equation[a]	t
Expected Use	.02	(1, 208)	4.46*	Evaluator Control	2.11*
Level of Impact	--	--	NS	--	--
Instrumental Use	.18	(2, 183)	20.65***	Stakeholder Diversity	5.47***
Conceptual Use (Substance)	.29	(3, 192)	16.42***	Stakeholder Diversity	3.37***
				Evaluator Control	2.96***
				Depth of Participation	2.09***
Conceptual Use (Process)	.30	(4, 190)	20.56***	Depth of Participation	3.76***
				Evaluator Control	3.47***
				Stakeholder Diversity	2.55**
				Impact Expectation	2.07*
Affective Experience	.05	(1, 213)	10.51***	Stakeholder Diversity	3.24***

[a] In order of step of entry

* $p < .05$; **; $p < .01$; ***; $p < .001$

4.4 DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS

The eight demographic / background characteristic variables used in prior analyses (chapter 2) were employed to structure a multivariate analysis of variance in order to determine whether collaborative evaluation practice depended upon such features. To repeat, the eight categorical variables were: response option (parts B + C, all parts, all parts plus practitioner nominated); nationality (American, Canadian); organizational location (5 possibilities); gender; years of experience as an evaluator (few, moderate, many); evaluation as prime responsibility (yes, no); percentage of evaluation projects involving practitioners (low, moderate, high); and training (doctorate, masters, other) (see chapter 2 for a more complete description). Analyses were run for three clusters of dependent variables corresponding to impact, and process issues. For each analysis, a single demographic / background variable was entered as the independent variable: no search for interaction effects was carried out.

(1) Impact

Multivariate effects on impact were attributable to evaluators' years of experience, $F(8, 562) = 1.99, p. < .09$; the percentage of evaluations conducted that were considered collaborative, $F(8, 546) = 2.94, p. < .01$; and evaluator gender $F(4, 287) = 2.38, p. < .05$. That is to say, at the univariate level, more seasoned evaluators tended to report higher levels of expected utilization, $F(2, 285) = 3.50, p. < .05$, and more favourable affective experiences with collaborative evaluation, $F(2, 285) = 3.73, p. < .05$. Evaluators who engage more frequently in collaborative evaluation reported higher levels of expected utilization, $F(2,277) = 3.30, p. < .05$; and more favourable affect as a result of their involvement in collaborative evaluation, $F(2, 277) = 5.20, p. < .01$. There was also a slight tendency (albeit not statistically significant) for males to derive higher levels of satisfaction from the process than females.

In two cases where multivariate effects were not observed, univariate F tests showed relationships between background characteristics and measures of impact. Expected utilization was found to be higher for evaluators who had evaluation as a primary responsibility (approaching significance) and for Canadian evaluators, $F(1, 293) = 6.36, p. < .01$.

(2) Collaborative Evaluation Process Variables

Three process variables were differentiated by multivariate effects of demographic / background variables: nationality, $F(5, 260) = 5.20, p < .01$; extent to which evaluators engage in collaborative evaluation, $F(10, 484) = 2.19, p. < .05$, and the response option selected for the present study. Canadians reported involving practitioners in a greater number of evaluation tasks than their American counterparts, $F(1, 293) = 6.36, p. < .01$. Those who frequently practice collaborative evaluation reported relying on a more diverse set of stakeholders for participation, $F(2, 247) = 6.85, p. < .001$ and to tend to maintain control of the evaluation decision making, $F(2, 247) = 3.02, p. < .05$. Finally, evaluators nominating a practitioner colleague reported working with a greater number of stakeholders, $F(1, 261) = 5.51, p. < .05$ and involving practitioners in a greater number of evaluation tasks, $F(1, 261) = 4.65, p. < .05$. Finally, evaluators with evaluation as their primary responsibility were more likely to report collaborative projects where they remained in control of evaluation decision making, $F(1, 259) = 4.61, p. < .05$.

4.5 SUMMARY

Evaluators tended to report favourable levels of impact of the collaborative evaluations they described, with some indication that impact extended beyond specific adherence to evaluation data. In particular, the stimulation of thinking about assumptions underlying practice may have occurred. Generally, evaluators appeared to have enjoyed the projects they worked on, although they may have found them stressful, frustrating and inefficient at times. Respondents also reported having disseminated findings to a wide array of stakeholder groups through a variety of modes. Regarding the collaborative evaluations described, evaluators tended to remain in

control of decision making and cast program practitioner involvement as being limited to scope definition and interpretive phases of the study. They also tended to involve a substantially diverse group of stakeholders in the project, although most relied heavily on primary users.

Relationships were noted between collaborative evaluation processes and impact. Evaluations involving and reporting to diverse groups of stakeholders tended to have more impact, but evaluators maintained a strong role in these projects. Evaluators who held favourable attitudes toward collaborative evaluation tended to report higher levels of use and personal satisfaction from the project they described. Finally, a variety of demographic / background characteristic variables served to differentiate reported impact and collaborative evaluation processes. The largest effects were attributable to evaluators who reported more experience with collaborative evaluation, although effects were also noted for Canadian evaluators, evaluators with more experience and those with more training.

CHAPTER 5: CONVERGENCE OF EVALUATOR AND PRACTITIONER PERSPECTIVES

Sixty-eight (58.6%) of the practitioners returned usable questionnaires. Since the version of the questionnaire for program practitioners was virtually a mirror image of that for the evaluators, the basis for comparison is good. The data permit answers to the following research questions.

R6 To what extent to researcher and practitioner views and opinions about collaborative evaluation converge? . . . perceptions about the process of the collaborative project and its impact? What are the nature of the differences in perception.

R7 Do demographic / background characteristics differentiate practitioners' descriptions and perceptions about the focal collaborative evaluation? . . . its processes and consequences?

Answers to these questions will be provided by first considering the descriptive characteristics of the practitioner responses and then examining the convergence between evaluators and practitioners of opinions and perceptions of practice. Finally, differences attributable to background characteristics of practitioners will be considered.

5.1 DESCRIPTIVE CHARACTERISTICS OF THE PRACTITIONER SAMPLE

Appendix A (page 8) shows descriptive statistics on practitioner background characteristics. A rather large proportion (almost half) of the sample were located in government agencies with 14% located in either non-government public sector organizations or private sector organizations that depend on public funds. Over 80% indicated that evaluation is common in their organization. Almost a 2:1 ratio of females to males was observed with respondents having about 17 years experience, on average, in their current field of expertise. Most of the respondents were American (60.3%) with the remainder being Canadian. On average, practitioners indicated that about three quarters of the evaluation projects carried in their home organizations involved practitioners. Practitioners were well educated with 65% holding a master's degree or higher.

5.2 VIEWS AND OPINIONS

Practitioners were asked to respond to the same Likert-type items given to the evaluators. The displays appearing in Appendix A are misleading to the extent that practitioner response summaries are juxtaposed to the entire evaluator sample, as opposed to the subsample of 68 evaluators who were paired with the practitioners. Nevertheless, it can be seen in the appendix that the item level magnitudes of the practitioner average ratings follow, within reason, those provided by the evaluator sample. Regarding general evaluation issues, practitioners were of the opinion, as were evaluators, that evaluators should be involved in follow-up activities and should

formulate recommendations for the study. They further believed that internal evaluation is becoming more well established and that evaluators should have as a primary function maximizing intended uses of data by intended users. Practitioners like evaluators, did not believe that summative evaluations should be limited to external evaluators. Also, they did not see the optimization of social justice as a central concern for evaluators.

Views about practitioner participation in formative evaluation also yielded a similar pattern to that of evaluators. Practitioners supported the idea of evaluators educating practitioners about the power of evaluation for planned change and they endorsed the perspective suggesting that evaluators can stimulate practitioners to think deeply about practice and to work toward changing practice. They supported the observation that practitioner participation makes evaluation more responsive to the needs of those using the data, and that it fosters higher levels of data utilization. However, practitioners, like evaluators, downgraded the proposal of having stakeholders with special interests being on board. Comparatively speaking, practitioners were not enamoured either with the idea that participation brings about social justice or that evaluator and practitioner control of the evaluation project should be shared (they supported evaluator control). However, ratings were slightly higher than evaluators on both of these issues items (items 2.11 and 2.13).

Respondents were provided with an opportunity to supply open-ended comments and 25 of the 68 took advantage. Twelve of the comments related to the specific collaborative evaluation that was the focus in Part A. Some examples were:

> The project was conducted essentially to meet compliance regulations. We knew going in that we would have an impact on year-round vs. traditional schooling -- that is, controlled by money and population factors. More research needs to be done on the actual calendar effect on student learning.
>
> I was very pleased with the way this participatory evaluation was carried out. Although a lot of work in a short period of time, it was fun and challenging. Had we evaluated performance, rather than process, it might have been more threatening, however.
>
> Participatory evaluation is a very enriching professional development, however, very time consuming. Those involved, underestimated time and energy involved. Should negotiate for time in advance. Some things suffered or had to be placed aside for the duration of study.

While these responses are generally favourable, they point to some significant concerns about the process and its fit within the normal goings-on of the organization. Other comments dealt with the interrelationships between the world of practice and the world of research:

> The qualitative, formative evaluation has especially been instrumental in involving practitioners in data driven change.
>
> Those responsible for evaluation should not beholden to those responsible for the product being evaluated.
>
> "Practitioners" to be effective must be prepared to design and implement concurrent evaluation of projects and programs. "Evaluators" can be helpful as consultants at various points during the implementation of the program, and as a summary evaluator at the conclusion of the program.
>
> Evaluators want change in program delivery and this can be done by altering evaluation practices. On the other hand, practitioners want evaluation tools that will work in terms of time and measuring student growth.

Finally, some respondents raised concerns about the awkward nature of some of the wording in the questionnaire. "Social justice" was raised as ambiguous by one respondent, as it was be several evaluators, and another found that many of the items required a concrete focus for a reliable responses to be given.

5.3 COLLABORATIVE EVALUATION PRACTICE

Practitioners also described in detail and shared their perceptions about the same evaluation project that served as the basis for their evaluator partner's responses. Again, the display in Appendix A will provide the reader with a quick overview of the response pattern for Part A of the questionnaire. Practitioners provided a similar estimate of project duration to that of evaluators (slightly over 1.5 years). Sixty-five respondents provided written comments in response to the open-ended item inquiring about the description of the evaluation project (purposes, methods, audience). Of those, 55 commented on purpose. Sixty-percent of those commenting indicated that the evaluation had a formative purpose. Some examples follow,

> To develop a process of program evaluation to be used to improve the quality of the cooperation program and for ensuring the attainment of Ministry policies.
>
> The purpose of the study is to determine if altering instructional practices, organizational design, and use of emerging technologies can effectively improve student learning in a middle school setting.
>
> Evaluation was intended to provide a means to ensure effectiveness of programs, to make mid-course corrections and impact.

> Design qualitative evaluation of a community development initiative to determine the effectiveness of CD approach (community mobilization, community-based assessment, intervention planning) in 15 ethnic communities in California.
>
> Grade 6 math review for entire system (50+ schools).

Other responses referred to evaluations that were decidedly judgment-, accountability-, or decision-oriented:

> To provide the governing body of the GEF with an assessment of the pilot phase of the GEF program (1991-93), preparatory to decisions on replenishment and restructuring.
>
> Purposes: To establish whether or not the project has been effective in increasing awareness about the 3 heart health risk factors and whether or not people have been influenced in their stage of change.
>
> Purpose was to ensure funders and state government that we are fulfilling our mission of sending well trained urban teachers into the classroom and that we are being as cost effective as possible.

A few of the comments indicated that the evaluation purposes were mixed. About 50% of the practitioners did not mention methods in their response. Of those who did, most (40.6%) indicated mixed-method approaches were used, while a substantial proportion (34%) identified exclusively qualitative methods such as interviews and document reviews. Twenty-five percent referred to exclusively quantitative designs. Finally, practitioners had a similar response pattern to evaluators regarding comments about intended audiences. Sixty percent of the responses contained information about audiences of which 61.5% listed program decision makers as the main recipient of the findings. More respondents than in the evaluator sample (17.9% compared with 9.4%) indicated program beneficiaries as recipients of evaluation outputs, while 20.5% suggested the main audience was a blend of people affiliated with the program and those for whom the program is intended.

Program practitioners had similar distribution of average ratings to evaluators concerning their involvement in the evaluation. Like evaluators they rated scope defining and interpretation / dissemination activities most highly and data processing and analysis activities much lower. Ratings for preparing reports and formulating recommendations appeared to be considerably higher that the evaluator sample. About 17 practitioners, on average, participated in the evaluation, although, as with the evaluator sample, the median of 6 is a more realistic summary of the group. Respondents indicated that practitioners tended to belong to more than one group, on average, and they rated, as did evaluators, program developers, managers and implementors as being most frequently involved. Beneficiaries and special interest groups were involved less

frequently. As did evaluators, practitioner respondents leaned slightly toward favouring evaluators concerning perceived control over the evaluation project, although a clear majority (73.5%) represented control as being balanced between researchers and practitioners.

Practitioners were almost evenly split between volunteering (42.1%) or being assigned (37.%) to the evaluation project, taking into account the 11% of responses that were not codable. For the remaining cases, mixed or ambiguous responses were supplied. Researchers became involved for a variety of reasons, according to the practitioners' perspective. Twenty-nine percent of the responses were too ambiguous to code, but of those that were more clear, the majority indicated that researchers were recruited for the project (63.6%). Some examples were, "Was hired to work with the group from the inception of the project.", "[The evaluator] was selected by State DOE Innovation co-ordinator."
"Evaluators were hired by the project to provide assistance in design, data collection, and analysis." A little more than 30% of the responses revealed that the evaluator was mandated or assigned to the activities: "Evaluation was mandated", "Assigned by donor to consult with the program" and "Assigned as task requirement for funding" were examples. These data contrast with those provided by the evaluator respondents in that the majority for that group fell into the "assigned" category, followed by "recruited" (reverse order from evaluators) Two respondents indicated that the evaluator joined the team on a volunteer basis, a proportion that is comparable to the other sample.

Practitioners had similar views to evaluators as to the role of evaluators on the project. Several of the 10 items listing various possibilities corresponded in magnitude between the two samples. The provision of guidance on technical matters, processing and analysing data, preparing reports, developing data collection instruments and formulating recommendations, were among the highest rated items on average, whereas chairing meetings and helping practitioners develop research skills were among the lowest.

Also, corresponding to evaluators perceptions were the average practitioner's views about the methodological orientation of the study and of dissemination and communication patterns. Practitioners tended to agree that the study was guided by a previously defined conceptual framework and their views about who evaluation results were communicated to also corresponded in large measure to evaluators'. Practitioners were somewhat more inclined to identify program beneficiaries. When asked how results were communicated, practitioners agreed with evaluators that usual modes of communication (executive summary, oral presentation, written report) were most often used. They also agreed that follow up meetings and newsletter-type communications were less frequently observed.

In their open-ended responses about impact practitioners had a great deal to say. Perhaps a little more so than evaluators, practitioner comments reflected significant impact. After removing the 14 cases who claimed it is too early to tell, almost 80% reported significant impact. This is slightly more than evaluators were inclined to do. Some examples were:

> The study indicated a need for better communication to families. This change has been effected.
>
> The study, thus far, is providing a model for change throughout all middle schools in the district.
>
> Funding for 95-96 was guided by the results of the study. Gateway is reprioritizing its efforts as a result of the study's findings.
>
> Major impact. Results have been systematically implemented . . .
>
> Used as basis of finding proposals (successful), used as basis of program development.

Some of the responses included views about what influenced outcomes.

> The study represented a milestone in the development of common elements for a performance framework. Implementation did not proceed because of profound changes in the organization.
>
> Implementation ongoing from study recommendations. New course development under way. Many road blocks due to lack of organization of original study framework, follow through of assistance to project.
>
> System used favourable responses from recipients to implement design elsewhere - state refused to further fund because evaluation did not clearly show success or impact.
>
> Funding for the program has continued. Program faculty are pursuing implementation of some of the recommendations.

Finally, about 22% wrote about limitations on observed impact due to a variety of reasons, including politics, aspects of the findings, communication problems and so on. Commenting on how the results were used, 34% of the sample did not provide adequate information. Of the remaining responses, over 80% wrote about improvement-oriented utilization of the data:

> The study, thus far, is providing a model for change throughout all middle schools in the district. The regional Ed service center (sponsor of study) is now extending this concept in a large scale program to all middle schools in our country.
>
> USDA make changes in program management and implementation of second series of demo projects. Individual projects made changes in project plans and implementation.

Continued implementation of project (multi-age program). Continued improvement of parental involvement through K - 3 parent/teacher conference and beginning of 4th/5th grade parent conferences. Continued and refining of leadership team to make decisions and work to improve instruction.

After one year, most recommendations seem on their way to be implemented.

The results have been used to restructure the club - enhancing our impact locally, but have not been applied to other state projects.

The remaining responses related to such matters as creating a new program, or generating additional resources or funding.

Impact measured quantitatively corresponded closely with the estimates provided by the evaluator sample. Average ratings to the set of items describing a range of consequences seemed to covary with the evaluator average ratings. Practitioners tended to support the observation that the evaluation caused them to think and reformulate their understanding of practical matters. Finally, ratings of dimensions of affect were invariably higher for practitioners than they were for evaluators. Clearly their involvement in evaluation was a rewarding experience even though some frustrations and concerns may have been evident.

5.4 TESTS FOR CONVERGENCE

As shown in the Appendix questionnaire items were grouped according to specific foci (e.g., evaluation tasks, perceived impact). For question sets, multivariate analysis of variance with repeated measures was carried out in order to test for differences between program practitioner and evaluator samples. Table 5.1 reports the observed differences. The multivariate test is reported for all question areas and for those tests that showed significant differences, univariate F tests were reported to reveal where the mean differences lay at the item level. In addition to the multivariate analyses, related t-tests were conducted for individual items and scales that were not grouped with others.

Respondents were asked to rate the extent of participation by practitioners on ten tasks associated with the evaluation process ranging from scoping out the study to disseminating its results. Generally, practitioners appeared to participate in roles that are consistent with the conventional stakeholder-based model. They helped to frame and shape the evaluation and to interpret and disseminate results, but played a minor role in the technical aspects of the study. Table 5.1 shows that practitioners tended to estimate their involvement in the study to be greater than the estimates provided by evaluators. This was particularly the case for the technical data processing and analyses tasks.

Table 5.1: Multivariate Analyses of Variance

ITEM CATEGORY / Differentiating Items	Researcher Mean	Researcher SD	Practitioner Mean	Practitioner SD	F	df	Prob.
PRACTITIONER PARTICIPATION (5 pt. frequency scale)	--	--	--	--	**3.04**	**10, 43**	**< .01**
Designing the study	3.50	1.13	3.87	1.07	3.43	1, 52	= .07
Processing for data analysis	2.15	1.23	2.81	1.43	10.74	1, 52	< .01
Data analysis	2.43	1.21	2.98	1.41	6.70	1, 52	< .01
STAKEHOLDER GROUPS INVOLVED (5 pt. frequency scale)	--	--	--	--	**NS**	--	--
EVALUATOR PARTICIPATION (5 pt. frequency scale)	--	--	--	--	**NS**	--	--
EVALUATION CONSEQUENCES (5 pt. frequency scale)	--	--	--	--	**5.20**	**7, 31**	**< .001**
Basis for significant decisions	3.79	0.84	3.56	0.82	3.45	1, 37	= .07
Development of research skills	3.00	0.90	2.55	1.03	5.47	1, 37	< .05
Basic assumptions questioned	3.50	0.95	3.90	0.89	4.46	1, 37	< .05
AFFECTIVE EXPERIENCE (5 pt. semantic differential)	--	--	--	--	**2.36**	**9, 53**	**< .05**
Stressful / not stressful	2.61	1.18	3.24	1.24	10.21	1, 61	< .01
Unmanagable / managable	3.89	0.96	4.29	0.69	10.09	1, 61	< .01
Frustrating / encouraging	3.66	1.17	4.06	0.98	4.10	1, 61	< .05
Rough / Smooth	3.32	0.95	3.77	0.84	8.83	1, 61	< .01

BEST COPY AVAILABLE

Table 5.1: Multivariate Analyses of Variance (continued)

ITEM CATEGORY / Differentiating Items	Researcher Mean	Researcher SD	Practitioner Mean	Practitioner SD	F	df	Prob.
OPINIONS ABOUT EVALUATION (4 pt. agreement scale)	--	--	--	--	--	--	--
VIEWS ON PARTICIPATION (4 pt. agreement scale)	--	--	--	--	**2.14**	**7, 41**	**= .06**
Involve primary users	3.27	0.57	3.04	0.62	3.43	1, 47	= .07
Involve special interest groups	2.70	0.61	2.40	0.71	4.97	1, 47	< .05
VIEWS ON PROFESSIONAL DEVELOPMENT (4 pt. agreement scale)	--	--	--	--	**2.74**	**5, 54**	**< .05**
Question fundamental beliefs	3.66	0.48	3.50	0.54	3.47	1, 58	= .07
Improve practice	3.78	0.41	3.51	0.57	8.49	1, 58	< .01
Educate re: power of evaluation	3.54	0.54	3.31	0.53	6.28	1, 58	< .01
VIEWS ON IMPACT (4 pt. agreement scale)	--	--	--	--	**NS**	--	--
VIEWS ON TECHNICAL QUALITY (4 pt. agreement scale)	--	--	--	--	**NS**	--	--
VIEWS ON EMPOWERMENT (4 pt. agreement scale)	--	--	--	--	**2.37**	**5, 45**	**< .05**
Educate re: power of evaluation	3.54	0.54	3.34	0.52	3.50	1, 49	= .07
Involve special interest groups	2.78	0.62	2.44	0.76	6.26	1, 49	< .05

BEST COPY AVAILABLE

Evaluators and practitioners agreed as to the number of practitioner participants who participated in the study, with the median being 6. They also concurred that practitioner participants belonged to more than one stakeholder group and a multivariate test comparing perceived involvement from six different categories of stakeholders--ranging from program developers to program beneficiaries and special interest groups--yielded no statistically significant differences.

The groups agreed about who controlled evaluation decision making with almost three quarters of the sample suggesting control be balanced while researcher control was viewed as greater than practitioner control slightly more often for the remaining cases. A series of ten items solicited respondents' views about the evaluator's role in the evaluation. Table 5.1 shows that no differences between practitioner and evaluator views were evident. Generally, evaluators were thought to be most highly involved in providing technical support and expertise and less involved in chairing project meetings, disseminating results, helping practitioners develop research skills and educating practitioners about the power and value of evaluation as a planned change strategy.

A variety of questionnaire items addressed the perceived impact of the collaborative evaluation project. While the respondent views converged about the perceived extent of impact as being modest with more expected, their views diverged about intended use by intended users. Researchers tended to provide higher estimates of intended uses than their practitioner counterparts, $t_{(60)} = 2.22$, $p. < .05$. On average, they believed that intended users were taking steps toward action based on the data, while practitioners were of the opinion that the information is still being considered. Ratings of seven different uses of findings and impacts of the evaluation process were analysed and revealed an interesting pattern (see Table 5.1). Evaluators tended to overestimate relative to practitioners the instrumental consequences of the evaluation. They also overestimated the role that the evaluation process played in developing among practitioner skills in doing research. However, practitioners were more inclined to suggest that working on the evaluation had caused them to question basic assumptions and beliefs about practice.

In rating their affective experience arising from the collaborative project, practitioners appeared to find involvement to be much more satisfying and enjoyable. Table 5.1 indicates that practitioners found the process to be less stressful and frustrating and more manageable and smooth than did their evaluator counterparts.

The next section of the questionnaire dealt with a variety of issues associated with evaluation in general, and particularly with formative, improvement-oriented evaluation. Opinions about evaluation in general, including evaluator roles, purposes and conduct, converged as shown in Table 5.1. Generally, these opinions were favourable. Respondents from both groups tended to disagree that the evaluator's role is to bring about social justice.

A list of 17 items were associated with views on formative, improvement-oriented evaluation. Factor analyses of these items reported in chapter 3 helped to categorize them into

five different categories and subsequent linear combinations were found to be internally consistent. The categories are views about: practitioner participation in evaluation; formative evaluation's role in fostering professional development; impact; issues associated with the technical quality of the evaluation; and formative evaluation's role in fostering empowerment among participants. Views on impact and technical quality converged over evaluator and program practitioner groups, while differences were noted for each of the other categories (see Table 5.1).

Practitioners were somewhat more skeptical than evaluators about the wisdom of involving primary users (individuals with a vital interest in the program and who are able to act on evaluation data) and stakeholders associated with special interest groups. They were also less inclined to acknowledge the potential for evaluation to enable practitioners to question fundamental beliefs about their practice, to improve their practice or to acknowledge the importance of the evaluator's role in educating program practitioners about the power and value of evaluation. Finally, they were less optimistic about the potential of evaluation to empower stakeholders.

5.5 EXPLAINING PATTERNS OF DIVERGENCE

The survey data reported above yield an interesting pattern with four distinct dimensions. First, evaluators tend to downplay the actual role that practitioners played in the technical evaluation tasks. Second, they also tended to overestimate the impact of the evaluation relative to their program practitioner counterparts. Interestingly, they underestimated the extent to which participation in the process caused practitioners to reflect on their practice. Third, the collaborative evaluation experience appeared to be somewhat more positive for practitioners. Comparatively speaking, evaluators struggled with the process of finding it stressful, difficult to manage and frustrating. Finally, attitudes about evaluation purposes, processes and roles were more favourable for evaluators than was the case for the practitioner respondent group.

As an aid to clarifying these divergences, we resort to the written comment data provided by respondents. Three open-ended questions solicited pertinent responses.

1. "Briefly describe the evaluation in terms of its (a) purposes, (b) methodological features and (c) main audience." The purpose of this question was to refresh the respondent's memory as she or he began the survey. Responses were provided prior to moving into the various ratings and questions on the specific project.

2. "What impact has the study had? How have the results been used? What factors affected (enhanced or limited) the impact of the study?" This question was sequenced after the respondent had rated various aspects of the evaluation process and prior to providing ratings on evaluation consequences.

3. "Additional comments about evaluation practice." This open-ended item appeared at the end of the section on views and opinions about evaluation.

A subset of respondent pairs were selected on the basis of their responses to numerical items associated with consequences of the evaluation. We selected the most discrepant pairs consistent with the pattern of mean differences. Six pairings (cases) were identified as being most different (usually they differed by two points on the five point scales) on the items in Table 5.1 under evaluation consequences and affective experiences. The direction of the discrepancies were consistent with the direction of the mean differences reported in that Table. The written responses to the three open-ended items for each of the high discrepancy pairings appear in Table 5.2.

Comments associated with project description were generally convergent between practitioners and evaluators. Purposes included accountability requirements, summative judgements about program functioning and accomplishments and program involvement. Typically multiple methods were used in the evaluations and the projects had multiple audiences.

Some divergences in opinion were illuminated by the comment data. In case 152, the evaluator reported numerous instrumental and improvement-oriented benefits of the evaluation, while the practitioner was comparatively less enthusiastic. In case 216 the evaluator commented on the "consciousness raising" consequence of the evaluation, a posture consistent with a self-determination agenda for program beneficiaries. The practitioner counterpart merely spoke of baseline information provision. The comments provided by the pairing in case 333 were consistent but the practitioner illuminated how the evaluation contributed to the questioning of fundamental assumptions being made about curriculum. In case 402 the evaluator sited several positive consequences of the evaluation and concluded with a comment about program cutbacks in times of fiscal restraint. For the practitioner this turn of events appears to have negated any perceived use of the data. In addition to acknowledging evaluation's role in program development, consistent with the evaluator, the case 543 practitioner included enhanced program marketing as an outcome. Finally, the pairing of case 551 converged in their opinions about how the evaluation provided a range of insights suitable for program development and strategic planning.

In the final open-ended question, some interesting differences emerged. In case 333, the evaluator referred to the challenges confronting empowerment evaluation but was optimistic about its worthiness in a more or less conventional, utilization-oriented way. The practitioner, on the other hand, commented about the impact of the evaluation process on her own professional development. In case 551 the evaluator stressed the need to integrate evaluation into program development activities, while the practitioner focused on a plethora of barriers that would limit this eventuality.

Table 5.2: Written Comment Data from Discrepant Cases

CASE ID	EVALUATOR COMMENTS	PRACTITIONER COMMENTS
152	**Project Description:** Evaluate a six-site school-based, yet community focused, substance abuse prevention project funded by the CA Dept of Alcohol Drug Programs. Methods include student surveys, behavior setting analyses, agency network analysis, interviews, and archival indicator analyses. **Impact:** This study is just drawing to a close and the final report is being written. Preliminary & Process Reports have been used to: support continued funding of projects, design new initiatives by the funding agency, enlarge projects at the six sites, improve participating agency buy-in at the six sites (they were mandated to develop conditions), and to test new research methods. Study impact was limited by the inexperience of project site directors. Also limited by staff changes at the funding agency. **Other:** Evaluation practice cannot be separated from program practice. The context of the project being evaluated is too often ignored. We write a history and process report early in the evaluation that helps describe the program context, environment, and structure.	**Project Description:** My understanding is the evaluation was for the state to update them on the status of projects in the state, survey and interview techniques were used as well as review of quarterly reports. **Impact:** I've used the study to start programs on site and for background information for report and studies. **Other:** None.

Table 5.2: Written Comment Data from Discrepant Cases (continued)

CASE ID	EVALUATOR COMMENTS	PRACTITIONER COMMENTS
216	**Project Description:** (a) determine producer adoption of rice practices; (b) personal interview, random sample, by co-op extension agents, evaluator-practitioner developed instrument, freq/means; (c) rice producers. **Impact:** Limited impact and use of results. Timetable for doing this internally/externally being developed as time of practitioner results. Process of involving practitioner has had beneficial effect in raising consciousness of evaluation's need and value. **Other:** None.	**Project Description:** The purpose was to determine the level of adoption of extension rice production recommendations by LA rice farmers. The survey was structured around "Key Indicators" which were previously used by quantity adoption of certain recommended practices. **Impact:** Establish baseline data on adoption of recommended practices. Provided information on rice producer's participation in extension programs. **Other:** None.
333	**Project Description:** Title: School Within School (SWS) at [name withheld] elementary school: The struggle over evaluation. An attempt to help teachers in different SWS conduct "improvement" evaluations of their own schools. In particular the SMILE school. **Impact:** So far, may have changed the teachers' view of reading needs of their students, caused them to purchase appropriate materials, gives them data for communicating with parents. **Other:** Empowerment eavluation catches professional evaluators in a mulitude of competing demands--but is challenging and potentially useful for improvement.	**Project Description:** To come up with an assessment tool that better meet our SWS (School Within School) goals and objectives; tool had 3 areas one-reading attitude survey, two-reading ?? analysis and three design a new repot card that reports meaningful information. **Impact:** The results of our assessment has made us aware of where our students are what gains they have made. It also has made as re-evaluate our curriculum and we also found a need to make changes in our assessment continuum of books. **Other:** I found being part of the program developing team as well as the assessment of the program rewarding. A lot of work, but definitely worth it. It helps to have outside resource such as Dr. [name withheld].

Table 5.2: Written Comment Data from Discrepant Cases (continued)

CASE ID	EVALUATOR COMMENTS	PRACTITIONER COMMENTS
402	**Project Description:** Review (formative but primary summative) evaluation of co-op education program for large school board. Surveys of teachers, students, parents, placement supervisors (business), review of records, interviews with administrators (e.g., vice-principals) and senior managers (e.g., superintendents). **Impact:** The study clearly affirmed the value of coop ed and identified obstacles to use, especially for advanced students. The report is still praised but the program management and scope has been reduced since the Board experienced severe fiscal restrictions and reduced programs that wre not mandated. Evaluation is only one consideration in difficult fiscal decisions. **Other:** I feel that the context and purposes shape the answers to many of your questions. One of my goals is "capacity building" but this is a developmental process and the evaluator must be expert in encouraging participation within the capabilities of the organization. "Over promising" is a major flaw in many evaluations that are not used.	**Project Description:** To develop a process of prog review to be used to improve the quality of the coop program and for ensuring the attainment of Ministry policies. Audience: educators, students, employers, administration. **Impact:** No because ministry mandates for coop changed. **Other:** None.
543	**Project Description:** In process. Evaluate a national training system in substance abuse prevention. Data collected from 10,000 participants' trainers and policy makers through surveys, interviews, observation, and document review. Multiple audiences. **Impact:** Impact/outcome data used in continued program development. **Other:** This utilization focused survey, neglects questions about the quality of "what" is being utilized.	**Project Description:** Evaluate nurse and mental health training programs. **Impact:** Helped to change curriculum design. Helped to market program for future funding. **Other:** None.

Table 5.2: Written Comment Data from Discrepant Cases (continued)

CASE ID	EVALUATOR COMMENTS	PRACTITIONER COMMENTS
551	**Project Description:** The heart health project of Brant, Haldimand, Norfold is a five year project with 10% of its annual budget allotted to evaluation. I have been a member of the evaluation committee for the past 4 years. Purpose of project is demonstration projects, therefore, evaluation is fundamental to each intervention. **Impact:** We are still in the project--it has been extended for an additional year. The evaluation strategies we have developed have been presented at OPHA and many Heart Health Conferences. We have had many negative and positive outcomes, some of our strategies have worked and some have been disasterous (but we have published these as well to help others avoid our mistakes). **Other:** Evaluation should be a fundamental component of what practitioners do. There needs to be an evaluation component to all activities or else what is the point in undertaking the process if you have no plan or anticipated outcome!!!	**Project Description:** (a) Purposes: to establish whether or not the project has been effective in increasing awareness about the 3 heart health risk factors and whether or not people have been influenced in their stage of change. (b) Methodological features: pre and post surveys, use of triangulation approaches, focus groups, stage of change, participant tracking. (c) Main audience: community residents. **Impact:** Definite impact with certain strategies employed, namely modelling and skill building strategies. How results used: used to justify which strategies are chosen as priority and which are chosen for diffusion purposes. Factors that affected impact of study. For eval purposes it was difficult to compare changes because of multiple variables affecting the audience--pre and post design was good but matched controls would have been better unfortunately this is too costly. **Other:** The most common barrier for practitioners is lack of access to resources, tools, or funding to conduct good quality evaluation (i.e., often it is too costly to get adequate sample sie, test, tools for reliability etc.) also could use technical help with study design for eval purposes.

5.6 DIFFERENTIATION BY DEMOGRAPHIC / BACKGROUND CHARACTERISTICS

Attitudinal and practice variables were subjected to between groups multivariate analysis of variance (restricted to practitioner sample only) in order to test whether background characteristics of the respondents made a difference. Seven independent variables were used for this analysis: nationality (Canadian, American); organizational location (5 possibilities); gender; years of experience in current field of specialization (few, moderate, many); familiarity of evaluation within the organization (low, moderate, high); percentage of evaluations involving practitioners within the organization (low, moderate, high); and training (doctorate, masters, other). Nationality, training and organizational location failed to yield multivariate or univariate effects. A multivariate effect of percentage of collaborative evaluations on impact was approaching significance. Within this set, expected utilization was found to be differentiated by this organizational characteristic, $F(2, 43) = 3.78$, $p < .05$. Oddly, a curvilinear relationship was observed with the mean for the moderate level exceeding both high and low.

Two other background characteristics were found to have univariate effects. First, gender differentiated attitudes toward collaborative evaluation as a way to improve the technical quality of an evaluation (approaching significance) and perceived control of the evaluation project, $F(1,47) = 8.60$, $p < .01$. In both cases males responded higher than females. Finally, attitudes toward technical quality of collaborative evaluation, $F(1,39) = 6.07$, $p < .05$ and expected utilization (approaching significance) were found to depend on whether or not evaluation was considered to be common within the organization, but in opposing ways. For practitioners from organizations where evaluation was more commonplace, utilization estimates were higher. For respondents from organizations that are not overly familiar with evaluation, an emphasis on the capability of collaborative evaluation influencing the technical quality of an evaluation was emphasized.

5.7 SUMMARY

Practitioner accounts of what happened during the target collaborative evaluation and their espoused attitudes toward this form of systematic inquiry corresponded very well with those provided by evaluators. However, multivariate analyses showed some points of departure between the two groups. Regarding practice, evaluators tended slightly to overestimate the diffusion of data and the impact of the study, whereas practitioners reported higher levels of personal satisfaction from participating in it. Comment data from practitioners tended to frame very clearly conceptual (learning) and improvement-oriented benefits of the evaluation. Utility was defined by them chiefly in this way as opposed to signalling, decision oriented or instrumental consequences. Background characteristics that had some differentiating effects appeared to be associated with the organization within which the practitioner worked. More favourable responses from the point of view of making a case for collaborative evaluation appeared to come from respondents whose organizations have institutionalized evaluation or at least introduced it into the culture.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

6.1 LIMITATIONS

At least four caveats need to be borne in mind as one considers the present data. First, the focus for the study is collaborative evaluation, an ill-defined, highly diverse term, used by different people in different ways. To accommodate this diversity, an explicit definition was not supplied to survey respondents. Rather, the term was loosely used to connote activities where evaluators (trained researchers) participate with practitioners (non-researchers) to carry out applied social research activities. It seems likely that this relatively unrestrictive perspective would accommodate quite a broad range of the approaches (such as those laid out in chapter 1) that can be located among the sectors of Figure 1.2.

It must be noted, however, that a great many evaluators simply do not do collaborative forms of evaluation. That is, for many, interaction with stakeholders is limited to consultatory activities to help define the parameters of the evaluation. This fact poses serious questions about obtaining a representative sample of evaluators who do collaborative evaluation. The present approach sampled major evaluation associations and to invite evaluator participation (i.e., fill out Part B only), regardless of whether they (a) do collaborative evaluation or (b) do evaluation at all. Indeed Shadish and Epstien (1987) showed that a large segment of the AEA population consists of evaluation scholars who do not consider themselves practitioners. Even with efforts to open the survey to all, one is left with the sense that members of the target sample who do not do or are not interested in the topic were reluctant to participate (and, in most cases, likely did not). Nonetheless, differences observed in the data attributable to response option suggest that respondents completing sections B and C only were less favourable toward the general approach. This suggests that the achieved sample steps at least somewhat beyond evaluators who might be considered aficionados of collaborative evaluation.

A second caveat is that evaluators who do collaborative evaluation were provided with the freedom to select a project of their own choosing as a focal point for reflection. It is likely that the project selected numbered among their perceived success stories. While the results are nevertheless useful and interesting, it would be equally interesting to examine in detail projects that flopped or in significant ways did not live up to expectations. The sampling procedure used was unlikely to produce many selections that would fall into this category.

Third, the approach used to obtain the practitioner sample did not leave much doubt that respondents would be "on board" with this general approach to evaluation. While an examination of the ways in which perceptions and perspectives between evaluators and practitioners differed, especially since they were expected to be like-minded about collaborative evaluation, such points of departure were largely absent in the present data.

Finally, questions about concrete practice (Part A) seemed much easier for participants to respond to. Section B on general views and opinions created a certain amount of consternation among some of the respondents who were uncomfortable trying to generalize statements out of context. This section of the instrument, while producing some interesting and interpretable response patterns, likely also produced unwarranted error of measurement; findings should be interpreted with caution.

6.2 IMPLICATIONS FOR THEORY AND RESEARCH

Having laid out its limitations, it is fair to say that the survey produced some very interesting trends and findings. The framework used for discussing them has been outlined in chapter 1. First, consideration is given to issues of interest or arguments for doing collaborative evaluation in light of what the present data have to say. Then, the dimensions of form outlined in chapter 1 are revisited from the standpoint of current practices and perceptions as shared by North American evaluators. Finally, other aspects of the findings will be discusses in terms of the research literature.

6.2.1 Interests

Three arguments for doing collaborative evaluation were presented in chapter 1. They were pragmatic interests, relating to enhancing impact of evaluation on local practice, policy and organizational development; philosophic interests, corresponding to improving the quality of data through grounding them more deeply in the world and cultures of practice; and political interests, an argument for using collaborative evaluation to help in ameliorating social inequities and bringing about societal reform. In a sense the deck was stacked in the present survey toward generating data corresponding to pragmatic as opposed to philosophic or political arguments. Such interests will be considered first.

Evaluators and practitioners alike favoured enhancing the intended use of data by intended users as being the evaluator's primary function. Indeed many of the reports on the impact of collaborative evaluation were testimony to such outcomes. The sampling limitations sited above hardly render this finding as being surprising. But two considerations about impact seem important to raise. First, much of the impact that was documented either qualitatively (written comments) or quantitatively (ratings) suggested that the most significant forms of impact were, at some level, conceptual in nature. That is to say, they favoured the improvement-oriented function of evaluation, as opposed to judgemental perspective highlighting evaluations' role in assessing program merit and worth. Second, much of what was reported tended to extend beyond program-specific consequences of the data. Rated highly among types of impact were observations that evaluation can be integrated as a planned change strategy and that it can stimulate deep thinking about practice, assumptions underlying it and, ultimately, changes in it. Such consequences may extend beyond the focal program and are consistent with contemporary

perspectives on the directions in which evaluation ought to be going (e.g., Cousins & Earl, 1995; Fetterman, 1994; Jenlink, 1994; Matheson, 1994; Preskill, 1994a).

Many of the approaches mentioned above advocate the use of collaborative evaluation for improvement-oriented evaluation problems. Evidence from the current study, however, suggests that many projects do indeed focus on summative, judgemental and decision-oriented evaluation problems, situations where bias of a self-serving nature has the potential to become problematic. Given that collaborative evaluation appears to be occurring with such frequency in the summative context it becomes increasingly important to study closely how bias enters into and affects the process and to what extent checks against it are both in place and effective.

There were limited data concerning philosophical arguments for collaborative evaluation. Evaluators' tended to be of the view that collaborative processes did not necessarily improve the technical quality of evaluation. In fact, it stands to reason that in some cases technical quality may have been traded off for responsiveness. Many of the evaluation designs employed were pre-ordinate in orientation. That is, a conceptual framework was developed and applied in advance of data collection. Arguments for responsiveness in evaluation would suggest that the quality of data would be enhanced by implementing less structured, more emergent and responsive designs (Guba & Lincoln, 1989; Stake, 1975, 1983). This did not appear to be the case among respondents in the present survey. A more direct empirical examination of the philosophical argument for collaborative evaluation would be helpful.

Finally, many evaluators, and indeed practitioners, did not adhere to the amelioration of social justice as a central concern of evaluation, or probably more often, did not explicitly connect such concerns to evaluation. In fairness, it must be noted that the term was not well defined, and, as House has clearly laid out, in and of itself, social justice is a very complex and diverse phenomenon. Nonetheless, many shared the view that such concerns were secondary to the more pragmatic function of evaluation. A lack of connectedness of issues of social justice and empowerment to evaluation practice has been observed previously (e.g., Patton, 1994), yet many evaluation scholars advocate change in this direction (Fetterman, 1994; Sirotnik, 1990). While it is fair to say that further empirical work connecting evaluation practice to political ideologies is needed, research methods will need to be suited to understanding the phenomena at hand. How, for example, would one go about operationalizing "illumination", and "liberation", central arguments in the case for empowerment evaluation (Fetterman, 1994)?

6.2.2 Form

Control of the evaluation process, stakeholder selection, and depth of participation were presented as three fundamental and independent dimensions of form relevant to discussions of collaborative evaluation. Figure 1.2 provides a three-dimensional structure on which various forms of collaborative evaluation can be located. Data from the present survey provide some insights into the how current practice falls out on these dimensions, in addition to evaluators

views about them. First, both evaluators and program practitioners believe in and implement evaluations in which they maintain substantial control. Cousins and Earl (1995) advocate shared control or equal partnership; the evaluator assuming responsibility for input on technical research matters and working to minimize biases and the practitioner laying claim to substantive and contextual issues as a way of informing the evaluation. Other forms of collaborative evaluation advocate practitioner control, and a facilitory role for evaluators (e.g., Fetterman, 1994). North American evaluators appear to lean more toward models with substantial interaction and exchange but where control for evaluation decision making is concerned, they were reluctant to part with it.

Coming from a pragmatic orientation, Cousins and Earl (1995) also support the view that stakeholder selection should be limited to primary users, those who have the organizational authority and influence to act on findings and who are connected to program management and direction. The present findings support this to the extent that such users were most frequently among those involved. But the findings also suggest that projects involving a broader range of stakeholders had a more substantial impact. This may be true in part because such projects also had wider dissemination patterns. Further research is needed to examine the implications of stakeholder diversity. For example, to what extent does consensus building among diverse groups intrude on the timeliness of evaluation data?

Finally, interesting findings emerged concerning depth of participation. Initially, Cousins and Earl (1992) were of the view that practitioner involvement in evaluation should extend to the entire evaluation process. Their thinking was aligned with a corollary of the pedagogical adage, "you never really know something until you have to teach it." For practitioners working on collaborative evaluations it would be "you never really know your data until you have to analyse and write it up." In their recent work, Cousins and Earl (1995) acknowledge that research is showing that such complete and comprehensive participation may not be necessary, and indeed, from a practical standpoint, prudent. Data from the present study show that stakeholder participation is generally limited to evaluation tasks that are not heavily technical, although certainly direct participation in instrument development and data collection, as was observed, is more deeply involved than merely informing the evaluation design and helping to interpret data. Depth of participation was not found in the present study to be predictive of impact.

6.2.3 Other Findings

It was observed in chapter 1 that although a body of empirical research was emerging, there is a need to look at the extent to which findings generalize. Apart from the foregoing, the present survey is able to do so for two issues. First, participant gratification, noted to be strong in other studies, was also observed to be apparent in the current study. Practitioners found their participation to be a very positive experience on average, more so than did evaluators. There was not much evidence in the written comments to suggest that their development of research skills was part of this positive experience. In fact, both practitioners and evaluators downplayed

practitioner skill development as an activity on the project. Second, some effects on researchers were observed. Researchers, too, find the process to be positive although somewhat less so than practitioners. They did report it to be somewhat stressful and limits on efficiency were noted. Nonetheless, several positive reports came from evaluators who had witnessed first-hand the sorts of impact that a collaborative project could have, in terms of shaping the evaluation, its dissemination and its impact.

6.3 CONCLUDING REMARKS

The present survey is intended to provide some insights into the nature and consequences of collaborative forms of evaluation occurring in North America, to illuminate evaluator's views and opinions about such matters and to check their perceptions against those of a practitioner colleague. The findings should be treated with some caution for reasons mentioned above. These data are unable to, nor are they intended to, come down on one side or the other about whether evaluators ought to embrace collaborative evaluation as a legitimate direction within the profession. They do, however, highlight some discrepancies between theory and practice (e.g., use for summative evaluation) and add to knowledge about what practice looks like and what sorts of impact might be expected. In many ways the data are positive in this regard, but further, consistent and systematic inquiry will be needed to better understand the conditions under which collaborative evaluation may be appropriate and how to improve its implementation. Such knowledge will serve well evaluation and program practitioners alike.

REFERENCES

Alkin, M. C. (1990). Debates on evaluation. Newbury Park, CA: Sage.

Alkin, M. C. (1991). Evaluation theory development: II. In M. W. McLaughlin, & D. C. Phillips (Eds.), Evaluation and Education: At Quarter Century (pp. 91-112). Chicago, IL: The University of Chicago Press.

Alkin, M. C., & Coyle, K. (1988). Thoughts on evaluation utilization, misutilization, and non-utilization. Studies in Educational Evaluation, 14, 331-340.

Alkin, M. C., & Stecher, B. (1983). Evaluation in context: Information use in elementary school decision making. Studies in Educational Evaluation, 9(1), 23-32.

Alvik, T. (1995). School-based evaluation: A close-up. Studies in Educational Evaluation, 21, 311-343.

Argyris, C. (1993). Knowledge for action: A guide to overcoming barriers to organizationl change. San Francisco, CA: Jossey-Bass.

Argyris, C., & Schon, D. A. (1991). Participatory action research and action science: A commentary. In W. F. Whyte (Ed.), Participatory action research (pp. 85-96). Newbury Park, CA: Sage.

Ayers, T. D. (1987). Stakeholders as partners in evaluation: A stakeholder-collaborative approach. Evaluation and Program Planning, 10, 263-271.

Bickel, W. E., & Cooley, W. W. (1985). Decision-oriented educational research in school districts: The role of dissemination processes. Studies in Educational Evaluation, 11, 183-203.

Bickel, W. E., & Hattrup, R. A. (1995). Teachers and researchers in collaboration: Reflections on the process. American Educational Research Journal, 32(1), 35-62.

Bragg, D. D. (1995). Collaborative evaluation: An unconventional approach to overcoming evaluation-avoidance. Mimeo. University of Illinois, Urbana-Champagne.

Brandon, P. R. (1994, November). Involving program beneficiaries in reviews of evaluators' recommendations for curriculum revisions. Paper presented at the American Evaluation Association Annual Meeting, Boston, MA.

Brunner, I., & Guzman, A. (1989). Participatory evaluation: A tool to assess projects and empower people. In R. F. Connor, & M. H. Hendricks (Eds.), New directions for program evaluation: International innovations in evaluation methodology. (pp. No. 42, 9-17). San Francisco: Jossey-Bass.

Bryk, A. (1983). Stakeholder-based evaluation. New Directions in Program Evaluation, No. 17. San Francisco: Jossey-Bass.

Carr, W., & Kemmis, S. (1992). Becoming critical: Education, knowledge and action research. London: Falmer.

Cochran-Smith, M., & Lytle, S. (1993). Inside/outside: Teacher research and knowledge. New York: Teachers College Press.

Corey, S. (1953). Action research to improve school practice. New York: Teachers College Press.

Cousins, J. B. (1995). Consequences of researcher involvement in participatory evaluation. Studies in Educational Evaluation, 22(1), 3-27.

Cousins, J. B. (1995a). Assessing program needs using participatory evaluation: A comparison of high and marginal success cases. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 55-71). London: Falmer.

Cousins, J. B. (1995b, April). Organizational conseqeunces of participatory evaluation: A longitudinal case study. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. Educational Evaluation and Policy Analysis, 14(4), 397-418.

Cousins, J. B., & Earl, L. M. (Eds.). (1995). Participatory evaluation in education: Studies in evaluation use and organizational learning. London: Falmer Press.

Cousins, J. B., & Leithwood, K. A. (1993). Enhancing knowledge utilization as a strategy for school improvement. Knowledge: Creation, Diffusion, Utilization, 14(3), 305-333.

Cousins, J. B., & Simon, M. (1996). The Nature and Impact of Policy-induced partnerships between research and practice communities. Educational Evaluation and Policy Analysis, 18(3), 199-218.

Cousins, J. B. & Walker, C. (1995, June). Personal teacher efficacy as a predictor of teachers' attitudes toward applied educational research. Paper presented at the annual meeting of the Canadian Association for the Study of Educational Administration, June, 1995, Montreal).

Earl, L. (1995). District-wide evaluation of school improvement: A system partners approach. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning. London: Falmer Press.

Elliot, J. (1991). Action research for educational change. Milton Keynes: Open University Press.

Fetterman, D. (1994). Empowerment evaluation. Evaluation Practice, 15(1), 1-15.

Fetterman, D. M. (1995). In response. Evaluation Practice, 16(2), 179-199.

Garaway, G. B. (1995). Participatory evaluation. Studies in Educational Evaluation, 21, 85-102.

Gitlin, A., Bringhurst, K., Burns, M., Cooley, V., Myers, B., Price, K., Russell, R., & Tiess, P. (1992). Teachers voices for school change. New York: Teachers College Press.

Green, L. W., George, M. A., Daniel, M., Frankish, C. J., Herbert, C. J., Bowie, W. R., O'Neill, M. (1995). Study of participatory research in health promotion. Vancouver, BC: Royal Society of Canada.

Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. Evaluation Review, 12(2), 91-116.

Greene, J. (1994, November). Specific practices that facilitate action. Paper presented at the at the annual meeting of American Evaluation Association, Boston.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

House, E. R. (1991). Evaluation and social justice: Where are we? In M. W. McLaughlin, & D. C. Phillips (Eds.), Evaluation and Education: At Quarter Century (pp. 233-247). Chicago, IL: The University of Chicago Press.

Huberman, M. (1990). Linkage between researchers and practitioners: A qualitative study. American Educational Research Journal, 27(2), 363-391.

Huberman, M. (1994). Resarch utilization: The state of the art. Knowledge and Policy, 7(4), 13-33.

Huberman, M. (1995). The many modes of participatory evaluation. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 103-111). London: Falmer.

Huberman, M., & Cox, P. (1990). Evaluation utilization: Building links between action and reflection. Studies in Educational Evaluation, 16, 157-179.

Hustler, D., Cassidy, T., & Cuff, T. E. (Eds.). (1986). Action research in classrooms and schools. London: Allen & Unwin.

Jenlink, P. M. (1994). Using evaluation to understand the learning architecture of an organization. Evaluation and Program Planning, 17(3), 315-325.

Johnson, R. B. (1995). Estimating an evaluation utilization model using conjoint measurement and analysis. Evaluation Review, 19(3), 313-338.

King, J. A. (1995). Involving practitioners in evaluation studies: How viable is collaborative evaluation in schools? In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 86-102). London: Falmer.

King, J. A., & Lonnquist, M. P. (1992, December). A review of writing on action research: 1944-present. Minneapolis, MN: Center for Applied Research and Educational Improvement.

Lafleur, C. (1995). A participatory approach to district-level program evaluation: The dynamics of internal evaluation. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 33-54). London: Falmer.

Larson, J. K. (1981). Knowledge utilization: Current issues. In R. F. Rich (Ed.), The knowledge cycle. Beverly Hills, CA: Sage.

Lee, L. E., & Cousins, J. B. (1995). Participation in evaluation of funded school improvement: Effects and supporting conditions. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 72-85). London: Falmer.

Levin, B. (1993). Collaborative research in and with organizations. Qualitative Studies in Education, 6(4), 331-340.

Lewin, K. (1946). Action research and minority problems. Journal of Social Issues, 2, 34-46.

Louis, K. S. (1995, May). Two steps forward, one step back: Do we need a new theory of dissemination and knowledge utilization? Paper presented at the symposium Dissemination and Utilization of Research Knowledge, Oslo, Norway.

Love, A. J. (1991). Internal evaluation: Building organizations from within. Newbury Park: Sage.

MacDonald, B. L. (1976). Evaluation and the control of education. In D. A. Tawney (Ed.), Curriculum evaluation today: Trends and implications. London: MacMillan (Schools Council Research Studies).

Mark, M., & Shotland, R. L. (1985). Stakeholder-based evaluation and value judgments. Evaluation Review,

Mathison, S. (1994). Rethinking the evaluator role: Partnerships between organizations and evaluators. Evaluation and Program Planning, 17(3), 299-304.

McTaggart, R. (1991a). Principles for participatory action research. Adult Education Quarterly, 41(3), 168-187.

McTaggart, R. (1991b). When democratic evaluation doesn't seem democratic. Evaluation Practice, 12(1), 9-21.

Mertens, D. M., Berkeley, T. R., & Lopez, S. D. (1995). Using participatory evaluation in an international context. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 140-156). London: Falmer.

Nevo, D. (1993). The evaluation minded school: An application of perceptions from program evaluation. Evaluation Practice, 14(1), 39-47.

Nevo, D. (1994) Combining internal and external evaluation: A case for school-based evaluation. Studies in Educational Evaluation, 20, 87-98.

Noffke, S. E. (1992). The work and workplace of teachers in action research. Teaching and Teacher Education, 8(1), 15-29.

Oja, S. N., & Smulyan, L. (1989). Collaborative action research: A developmental approach. London: Falmer Press.

Owen, J. M., & Lambert, F. C. (1994, November). Evaluation for program clarification: An emerging utilization imperative. Paper presented at the American Evaluation Association Annual Meeting, Boston, MA.

Parlett, M., & Dearden, G. (Eds.). (1977). Introduction to illuminative evaluation: Studies in higher education. Cardiff-by- the-sea: Pacific Sounding Press.

Parlett, M., & Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovatory programs. From G. Glass (Ed/.), Evaluation Studies Review Annual Volume 1. Beverly Hills: Sage Publications.

Patton, M. Q. (1988a). The evaluator's responsibility for utilization. Evaluation Practice, 9(2), 5-24.

Patton, M. Q. (1986). Utilization-focused evaluation (2nd ed.). Newbury Park, CA: Sage.

Patton, M. Q. (1994). Developmental evaluation. Evaluation Practice, 15(3), 311-319.

Preskill, H. (1994a). Evaluation's role in enhancing organizational learning. Evaluation and Program Planning, 17(3), 291-297.

Preskill, H. (1994b, November). Facilitating organizational change: The use of focus groups to promote team and organizationa learning. Paper presented at the American Evaluation Association Annual Meeting, Boston, MA.

Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin, & D. C. Phillips (Eds.), Evaluation and Education: At Quarter Century (pp. 19-64). Chicago, IL: The University of Chicago Press.

Shadish, W. R., & Epstein, R. A. (1987). Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. Evaluation Review, 11(5), 555- 590.

Shapiro, J. P. (1988). Participatory evaluation: Towards a transformation of assessment for women's studies programs and projects. Educational Evaluation and Policy Analysis, 10(3), 191- 199.

Shulha, L., & Cousins, J. B. (1995, November). Utilization and social justice: Interconnections of meta-evaluation frameworks. Paper presented at the joint meeting of the Canadian Evaluation Society and the American Evaluation Association, Vancouver.

Shulha, L. M., & Wilson, R. J. (1995). Inviting collaboration: Insights into researcher-school community partnerships. In J. B. Cousins, & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 115-139). London: Falmer.

Sirotnik, K. A., & Oakes, J. (1990). Evaluation as critical inquiry: School improvement as a case in point. In K. A. Sirotnik (Ed.), Evaluation and social justice: Issues in public education. New directions in program evaluation. No. 45 (pp. 37- 59). San Francisco: Jossey-Bass.

Smith, N. L. (1983). Citizen involvement in evaluation: Emprical studies. Studies in Educational Evaluation, 9, 105-117.

Smith, N. L. (1980). Evaluation utility and client involvement in accreditation studies. Educational Evaluation and Policy Analysis, 2, 57-65.

Stake, R. E. (1983). Program evaluation, particularly responsive evaluation. In G. Madaus, M. Scriven, & D. Stufflebeam (Eds.), Evaluation models (pp. 291-300). Boston MA: Kluwer Academic Publishers.

Stake, R. E. (1975). To evaluate an arts program. In R. E. Stake (Ed.), Evaluating arts education: A responsive approach. Columbus OH: Charles E. Merrill.

Stufflebeam, D. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. Evaluation Practice, 15(3), 321-338.

Taba, H., & Noel, E. (1957). Action research: A case study. Washington DC: Association for Supervision and Curriculum Development.

Torres, R. T. (1994). Concluding remarks: Evaluation and learning organizations: Where do we go from here? Evalation and Program Planning, 17(3), 339-340.

Tripp, D. H. (1990). Socially critical action research. Theory into Practice, 29(3), 158-166.

Weiss, C. H. (1991). Evaluation research in the political context: Sixteen years and four administrations later. In M. W. McLaughlin, & D. C. Phillips (Eds.), Evaluation and Education: At Quarter Century (pp. 211-232). Chicago, IL: The University of Chicago Press.

Weiss, C. H. (1983). The stakeholder approach to evaluation. In A. Bryk (Ed.), Stakeholder-based evaluation, New Directions in Program Evaluation, No. 17 (pp. 3-14). San Francisco: Jossey- Bass.

Whyte, W. F. (Ed.). (1991). Participatory action research. Newbury Park: Sage.

APPENDIX A

ASS'N	EVALUATOR (N=564)	PRACTITIONER (N=68)
AEA	306	NA
CSE	200	NA
AERA	35	NA
AERO	23	NA
RESP OPTION		
B & C only	216	NA
A, B & C	232	68
A, B & C + Nom.	116	NA

Université d'Ottawa · University of Ottawa

SURVEY ON EVALUATION PRACTICE

This questionnaire is divided into three sections. Part A asks you to identify and respond to questions about a specific evaluation project on which you recently worked. Part B asks for your general views and opinions about evaluation. Part C asks for a minimal amount of background information about yourself.

Please complete and return the questionnaire in the self-addressed envelope (stamp provided) within two weeks of receipt. There is no need to put your name on the questionnaire. If you require further information about the survey please do not hesitate to contact us.

Dr. Brad Cousins, Associate Professor
Faculty of Education, University of Ottawa
145 Jean-Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5

Phone: (613) 562-5800 ext. 4088
Fax: (613) 562-5146
E-mail: bcousins@educ-1.edu.uottawa.ca

PART A: SPECIFIC EVALUATION PROJECT

In this section we are interested in your views about a specific evaluation project completed by you within the past three years. We are especially interested in applied social research (henceforth evaluation) studies in which practitioners who do not routinely conduct research as part of their role participate in carrying out the research. Please select such a study as the basis for your responses in Part A. **(If you have not carried out an evaluation of this sort check ✓ here _____ and GO TO PART B, PAGE 6)**

	Eval (217)			Prac. (NA)		
	x	SD	N	x	SD	N
months	18.1	19.0	303	19.1	17.4	53

1. When did the study (a) begin? _____/_____ (mo/yr) (b) finish _____/_____ (mo/yr)

2. Briefly describe the evaluation in terms of its (a) purposes, (b) methodological features and (c) main audience. ***(print or write legibly)***

_____

_____

_____

_____

3. In what ways did practitioners or non-researchers (henceforth practitioners) participate in the study?

For each of the following indicate by circling ONE option: N=Never; R=Rarely; S=Somewhat; F=Frequently; A=Always. Try to use N/A=Not Applicable as infrequently as possible.

3.1	Defining the scope of the investigation.	N	R	S	F	A	| N/A	4.01	0.10	344	4.09	1.15 66
3.2	Designing the study.	N	R	S	F	A	| N/A	3.35	1.17	344	3.86	1.08 66

		Evaluator			Practitioner		
		x	SD	N	x	SD	N
3.3	Developing data collection instruments. N R S F A | N/A	3.19	1.26	337	3.55	1.25	67
3.4	Collecting data/information. N R S F A | N/A	3.33	1.43	341	3.36	1.47	67
3.5	Processing and preparing data for analysis. N R S F A | N/A	2.14	1.33	338	2.70	1.43	67
3.6	Analyzing data. N R S F A | N/A	2.17	1.34	339	2.78	1.41	67
3.7	Interpreting results. N R S F A | N/A	3.23	1.34	339	3.33	1.38	67
3.8	Preparing reports for dissemination. N R S F A | N/A	2.72	1.42	337	3.20	1.43	66
3.9	Formulating recommendations from the study. N R S F A | N/A	3.35	1.40	331	3.86	1.24	66
3.10	Disseminating results and recommendations to intended users or audiences. N R S F A | N/A	3.68	1.23	326	3.97	1.19	62
4.	How many practitioners helped to carry out the evaluation? _____	19.4	52.3	326	16.9	44.38	64

		% Yes	No	N	% Yes	No	N
5a.	Did the practitioners who participated in the evaluation belong to more than one stakeholder group? _____ yes _____no	65.6	34.4	334	66.1	33.9	62
b.	Provide details _____ _____ _____						

6. Indicate the extent to which members of the following stakeholder groups helped to carry out the evaluation.

For each of the following, indicate by circling ONE option: N=Never; R=Rarely; S=Sometimes; F=Frequently; A=Always. Use N/A=Not Applicable (or Don't Know) as infrequently as possible.

		Evaluator			Practitioner		
		x	SD	N	x	SD	N
6.1	Program developers. N R S F A | N/A	3.48	1.32	305	3.72	1.20	61
6.2	Program managers or directors. N R S F A | N/A	3.75	1.09	327	3.86	1.02	64
6.3	Program sponsors or funders. N R S F A | N/A	2.65	1.39	285	2.78	1.39	58
6.4	Staff responsible for implementing the program. N R S F A | N/A	3.74	1.12	328	3.87	1.09	67
6.5	Intended beneficiaries of the program. N R S F A | N/A	2.65	1.33	320	2.71	1.22	62
6.6	Special interest groups. N R S F A | N/A	2.13	1.27	266	2.27	1.26	56
6.7	Other *(specify)* _____ N R S F A | N/A	3.76	0.96	28	3.54	1.05	13

BEST COPY AVAILABLE

7. How would you characterize the control of the evaluation project decision making process? *Check (✓) ONE option only*

_____ researcher controlled _____ practitioner controlled _____ shared / balanced

	Evaluator				Practitioner			
%	RC	PC	S/B	N	RC	PC	S/B	N
	27.3	9.2	63.5	337	14.7	73.5	11.8	68

8. How did practitioners come to be involved in the project (e.g., volunteered, task assigned by superiors)? *Differentiate by stakeholder group if applicable.*

_____

_____

_____

_____

9. How did you come to be involved in the project? Describe how your relationship with the practitioner group came about.

_____

_____

_____

_____

10. To what extent did you engage in the following activities during the project?
For Practitioners: "did the evaluator... ?"
For each of the following, indicate by circling ONE option: N=Never; R=Rarely; S=Sometimes; F=Frequently; A=Always. Use N/A=Not Applicable (or Don't Know) as infrequently as possible.

			Evaluator			Practitioner		
			x	SD	N	x	SD	N
10.1	Chairing project meetings.	N R S F A | N/A	3.26	1.44	325	2.99	1.42	68
10.2	Providing guidance about technical research matters.	N R S F A | N/A	4.92	0.83	339	4.16	0.84	68
10.3	Developing data collection instruments.	N R S F A | N/A	4.27	0.98	340	4.11	1.00	68
10.4	Collecting data.	N R S F A | N/A	3.73	1.39	339	3.56	1.42	68
10.5	Processing and analyzing data.	N R S F A | N/A	4.36	0.99	340	4.40	1.05	68
10.6	Preparing reports for dissemination.	N R S F A | N/A	4.89	0.84	338	4.34	1.08	67
10.7	Formulating recommendations from the study.	N R S F A | N/A	4.20	1.03	338	4.13	1.03	67
10.8	Disseminating results to intended users or audiences.	N R S F A | N/A	3.80	1.22	334	3.35	1.28	65
10.9	Helping practitioners to develop technical research skills.	N R S F A | N/A	3.11	1.33	321	3.15	1.24	66

BEST COPY AVAILABLE

	Evaluator x SD N	Practitioner x SD N
10.10 Educating practitioners about the power and value of evaluation as a planned change strategy. N R S F A | N/A	3.59 1.26 324	3.53 1.27 64
10.11 Other *(specify)* _____ N R S F A | N/A	4.42 0.89 31	4.67 0.58 3

11. Which of the following best describes your evaluation study? *(Check ONE option only)*	% N=339	% N=68
_____ the study was guided by a conceptual framework that was specified in advance of and guided data collection.	70.2	85.3
_____ a conceptual framework was constructed during and/or following data collection.	21.5	11.8
_____ the study did not make use of a conceptual framework.	8.3	2.9

12. To whom were the results/recommendations of the evaluation communicated? *(Check as many as apply.)*	% N=337	% N=66
_____ program developers _____ intended program beneficiaries	73.1 / 43.3	83.3 / 60.6
_____ program managers or directors _____ special interest groups	90.8 / 39.8	92.4 / 43.9
_____ program sponsors or funders _____ academic audiences	75.7 / 41.2	74.2 / 48.5
_____ implementors of the program _____ other *(specify)* _____	22.3 / 21.8	92.4 / 22.7

13. In what ways were results/recommendations of the evaluation communicated? *(Check as many as apply.)*	% N=337	% N=68
_____ executive summary _____ oral presentation(s)	81.6 / 80.1	77.9 / 77.9
_____ technical written report _____ follow-up committee meeting(s)	89.3 / 48.4	83.8 / 47.1
_____ newsletter/communication circular _____ other *(specify)* _____	24.6 / 23.4	30.9 / 22.1

14. What impact has the study had? How have the results been used? What factors affected (enhanced or limited) the impact of the study?

_____

_____

_____

_____

_____

_____

BEST COPY AVAILABLE

	Evaluator	Practitioner
15. Indicate which of the following best describes the impact of the evaluation? *Check ONE option only.*	% N=340	% N=68
_____ Has had an impact.	10.3	2.9
_____ Has had some impact, but is expected to have more.	39.4	42.6
_____ Is expected to have an impact, but it is too early to tell.	31.5	38.2
_____ Has not had an impact.	16.2	14.7
_____ Other (specify)_____	2.6	1.5

	% N=326	% N=65
16. Select the option that best describes <u>your opinion</u> about how information generated by the project was or is being used in general by the intended or expected "users." *Be sure to read each option carefully before checking ONE option only.*		
_____ It is premature for intended users to have knowledge of project information.	7.1	9.2
_____ Most intended users are not aware of information generated by the project.	7.4	7.7
_____ Most intended users are aware of information generated by the project through informal means only (e.g., word of mouth).	10.1	13.8
_____ Most intended users are currently considering information generated by the project (e.g., discussing with colleagues/peers).	20.2	23.1
_____ Based on the information, most intended users have taken steps toward action (e.g., decision to use, plans being made).	20.9	16.9
_____ Most intended users have taken action as a consequence of the information (e.g., made decisions, carried out plans).	34.4	29.2

17. Indicate the extent to which the following are (or will be) consequences of information generated by the project for intended users.

For each of the following, indicate by circling ONE option: N=Never; R=Rarely; S=Sometimes; F=Frequently; A=Always. Use N/A=Not Applicable (or Don't Know) as infrequently as possible.

		x	SD	N	x	SD	N
17.1	Intended users have based (or will base) significant decisions on this information. N R S F A | N/A	3.70	0.93	341	3.61	0.83	61
17.2	Intended users have learned (or will learn) about their practice. N R S F A | N/A	3.76	0.88	308	3.67	0.75	61
17.3	Intended users have developed (or will develop) their research skills. N R S F A | N/A	2.60	1.02	293	2.58	1.03	55
17.4	Intended users have learned (or will learn) that evaluation can be a powerful and valuable planned change strategy. N R S F A | N/A	3.46	0.98	319	3.73	0.88	63

BEST COPY AVAILABLE

			Evaluator			Practitioner		
			x	SD	N	x	SD	N
17.5	Data have helped (or will help) intended users incrementally improve their performance.	N R S F A | N/A	3.44	0.97	300	3.43	0.96	60
17.6	Data have helped (or will help) intended users to question basic assumptions and beliefs about their practice.	N R S F A | N/A	3.43	0.94	312	3.70	0.96	63
17.7	Data have helped (or will help) stimulate fundamental changes in practice.	N R S F A | N/A	3.24	1.03	312	3.61	0.97	61
18.	Describe your experience on this project. *For each of the following check (✓) ONE space only.*							
	Unfavorable ____ | ____ | ____ | ____ | ____ Favorable		4.27	0.92	340	4.49	0.75	65
	Unrewarding ____ | ____ | ____ | ____ | ____ Rewarding		4.22	0.95	339	4.35	0.76	65
	Stressful ____ | ____ | ____ | ____ | ____ Not stressful		2.67	1.23	338	3.20	1.24	65
	Unmanageable ____ | ____ | ____ | ____ | ____ Manageable		3.86	0.97	339	4.26	0.74	65
	Unpleasant ____ | ____ | ____ | ____ | ____ Pleasant		3.95	0.98	337	4.22	0.78	65
	Inefficient ____ | ____ | ____ | ____ | ____ Efficient		3.58	1.00	335	3.80	0.95	64
	Frustrating ____ | ____ | ____ | ____ | ____ Encouraging		3.43	1.21	339	3.97	1.05	65
	Ineffective ____ | ____ | ____ | ____ | ____ Effective		3.91	1.01	337	4.13	0.95	65
	Rough ____ | ____ | ____ | ____ | ____ Smooth		3.27	0.99	337	3.71	0.90	65

PART B VIEWS AND OPINIONS

For each of the following indicate by circling ONE option: SD=Strongly Disagree; D=Disagree; A=Agree; SA=Strongly Agree. Try to use N/A=Not Applicable as infrequently as possible.

1. General perspectives on evaluation.

1.1	The evaluator's primary function is to maximize intended uses by intended users of evaluation data.	SD D A SA | N/A	2.99	0.85	560	2.85	0.71	60
1.2	More and more organizations are establishing internal evaluation capabilities.	SD D A SA | N/A	2.84	0.81	516	3.05	0.63	59
1.3	Summative evaluations must be conducted by people external to the organization.	SD D A SA | N/A	2.32	0.91	554	2.17	0.82	64
1.4	Formative evaluations are best done by internal members of the organization.	SD D A SA | N/A	2.48	0.92	547	2.78	0.79	63

BEST COPY AVAILABLE

				Evaluator			Practitioner		
				x	SD	N	x	SD	N
1.5	The evaluator's primary function is to maximize the technical quality of the evaluation.	SD D A SA	N/A	2.75	0.91	549	2.80	0.83	66
1.6	Evaluators should be significantly involved in evaluation follow-up activities.	SD D A SA	N/A	2.96	0.76	542	3.14	1.05	66
1.7	The evaluator's primary function is to maximize opportunities to bring about social justice.	SD D A SA	N/A	2.15	0.85	506	2.18	1.27	61
1.8	Formative and summative evaluations cannot be separated.	SD D A SA	N/A	2.41	0.76	527	2.70	0.70	60
1.9	Evaluators should formulate recommendations from the study.	SD D A SA	N/A	3.20	0.72	545	3.16	1.04	67
1.10	Evaluators should have substantial expertise specific to the program being evaluated.	SD D A SA	N/A	2.50	0.80	554	2.88	1.06	66
	Views about practitioner participation in <u>formative</u> evaluation.								
2.1	The more stakeholder groups involved in evaluation the better.	SD D A SA	N/A	2.94	0.74	547	2.92	0.69	65
2.2	Involving multiple stakeholder groups helps to offset political agendas.	SD D A SA	N/A	2.93	0.76	533	3.05	0.67	61
2.3	Program beneficiaries should participate in carrying out evaluation.	SD D A SA	N/A	2.94	0.81	544	3.06	0.70	66
2.4	Special interest groups should participate in carrying out evaluation.	SD D A SA	N/A	2.55	0.77	510	2.45	0.75	60
2.5	People with a vital interest in programs (e.g., program developers, sponsors, directors) should participate in carrying out evaluations.	SD D A SA	N/A	3.01	0.79	544	3.06	0.61	66
2.6	People responsible for implementing or delivering programs should participate in carrying out evaluations.	SD D A SA	N/A	3.15	0.76	547	3.25	0.56	65
2.7	Practitioners' participation in evaluation makes evaluations more efficient.	SD D A SA	N/A	2.76	1.03	533	3.19	0.66	64
2.8	Practitioners' participation in evaluation makes the research more responsive to local needs.	SD D A SA	N/A	3.18	0.69	538	3.32	0.62	65
2.9	Practitioners' participation in evaluation enhances the technical quality of evaluations.	SD D A SA	N/A	2.52	1.05	529	2.86	0.72	62
2.10	Practitioners' participation in evaluation enhances the utilization of evaluation data.	SD D A SA	N/A	3.37	0.63	535	3.45	0.59	67
2.11	Practitioners' participation in evaluation helps to bring about social justice.	SD D A SA	N/A	2.56	1.09	404	2.75	1.30	48

BEST COPY AVAILABLE

		Evaluator x SD N	Practitioner x SD N
2.12	Evaluators should help train practitioners to do evaluation. SD D A SA | N/A	3.10 0.72 545	3.25 0.97 64
2.13	Evaluators should share control of evaluation projects equally with practitioners. SD D A SA | N/A	2.59 0.96 526	2.79 0.77 63
2.14	Evaluators should educate practitioners about the power and value of evaluation as a planned change strategy. SD D A SA | N/A	3.39 0.62 545	3.31 0.53 65
2.15	Evaluation can help practitioners improve practice. SD D A SA | N/A	3.58 0.55 551	3.54 0.56 67
2.16	Evaluation can help stimulate practitioners to question fundamental beliefs and assumptions about practice. SD D A SA | N/A	3.50 0.53 549	3.49 0.53 67
2.17	Evaluation can result in fundamental changes in practice. SD D A SA | N/A	3.45 0.58 547	3.49 0.59 66

3. Additional comments about evaluation practice:

_____

_____

_____

_____

_____

PART C: BACKGROUND INFORMATION

1. Your organizational location. *Check (✓) one option only.*

	% N=563	% N=66
_____ university (practitioners)	27.4 / 5.3	48.5 / 13.6
_____ private consulting firm (government)	12.3 / 5.9	13.6 / 4.5
_____ public sector organization (non-government)	41.7 / 7.5	3.0 / 16.7
_____ private sector organization (non-university)		
_____ self employed / freelance (non-consulting firm) (depending heavily on public funds)		
_____ other *(specify)* _____ (self-employed) (other / private sector)		

2. Your gender _____ female _____ male

% F M N	% F M N
49.3 50.4 560	65.7 34.3 67

3. Your years of experience as an evaluator or researcher _____ *(years)*
   For Practitioners: "in your current field of expertise."

X SD N	X SD N
13.6 8.03 543	16.6 8.84 67

4. Is evaluation your primary responsibility? _____yes _____no
   For Practitioners: "common in your organization?"

% Yes No N	% Yes No N
49.4 50.6 555	82.8 17.2 64

5. What percentage of evaluation projects that you have worked on have involved practitioners in carrying out the study?
   _____%

X SD N	X SD N
69.6 31.3 473	72.8 32.2 52

6. Highest degree obtained.
   Doctorate_____ Masters_____ Other_____ *(specify)*_____

% D M O N	% D M O N
55.1 38.0 7.0 561	15.2 51.5 33.3 66

THANK YOU FOR YOUR INPUT AND YOUR TIME

BEST COPY AVAILABLE

APPENDIX B1

Université d'Ottawa · University of Ottawa

Faculté d'éducation Faculty of Education

March 8, 1995

Dear Colleague

As part of a research and development project funded by the Social Sciences and Humanities Research Council of Canada, we are asking for your voluntary participation in a survey concerning program evaluation practices. In particular, we are interested in knowing more about the practices and perceptions of individuals such as yourself who routinely work on applied social research projects as evaluators/researchers. We are especially interested in your work on collaborative evaluation projects that involve members of the community of practice (e.g., practitioners, decision makers, intended users of evaluation data). The purpose of this research project is to learn more about the nature of such collaborative projects, influences on them and the impact that they have.

Your name was selected at random from the current membership list of the **AMERICAN EVALUATION ASSOCIATION (AEA)**. Administrative officials of the **AEA** have given us permission to use of the membership list for this purpose. You may be interested to know that members of other evaluation associations in Canada and the United States are also participating in the survey.

If you agree to participate, please take 20-30 minutes to fill out the attached questionnaire. Then follow one of the two options listed below. While it is our hope that you select option 2 please understand that your choice of at least option 1 is very important to us.

1. **Anonymous option**: Simply fold your completed questionnaire and return it within two weeks of receipt in the self-addressed envelope (stamp provided). If you would like to receive a two-page summary of the survey results you could (a) insert a note to that effect into the envelope OR (b) make your request known to the principal investigator (please find address information below). In any case, for all intents and purposes, your questionnaire responses will remain **ANONYMOUS**.

2. **Confidential option**: Part A of the questionnaire asks you a series of questions about a specific collaborative applied research project that you worked on recently. We would also like to hear from a practitioner who was involved in that project (i.e., participated in planning, carrying out, interpreting or reporting the study). Please select the practitioner who was most involved in the research project or is representative of those involved.

   The person that you identify will receive the attached (sample) letter and a questionnaire similar to the one that you fill out. All responses will remain strictly **CONFIDENTIAL**. Only members of the immediate research team will have access to individual questionnaire responses. These responses will be pooled with others for analysis. A master list of code numbers and names will be created strictly for the purpose of linking for comparison evaluator and practitioner responses. This list will be destroyed once all data have been entered onto the computer. Your name and address will not be used for any other purpose.

If you select option 2 please:

a. print the selected person's name and complete mailing address here:

_____

_____

_____

_____

b. print your name and organizational affiliation here:

_____

_____

c. provide an abbreviated title of the evaluation project that will be recognizable or familiar to the person you selected.

_____

d. return this letter **WITH** the completed questionnaire within two weeks of receipt in the self-addressed stamped envelope provided.

If you have inquiries about the project **or** wish to receive a two page summary of the results please contact the principal investigator at:

Mail Dr. Brad Cousins
Faculty of Education
University of Ottawa
145 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5

Phone (613) 562-5800 Ext. 4088

Fax (613) 562-5146

E-mail bcousins@educ-1.edu.uottawa.ca

Your participation is greatly appreciated.

Sincerely,

Brad Cousins
Associate Professor

APPENDIX B2

Université d'Ottawa · University of Ottawa

Faculté d'éducation Faculty of Education

SAMPLE

March, 1995

Dear Colleague

As part of a research and development project funded by the Social Sciences and Humanities Research Council of Canada, we are asking for your voluntary participation in a survey concerning program evaluation practices. We are especially interested in applied research projects that have the collaborative involvement of individuals such as yourself who do not routinely do research. The purpose of this research project is to learn more about the nature of such collaborative projects, influences on them and the impact that they have. To that end, we have already surveyed evaluators and researchers and one such person:

Name of researcher, organizational affiliation, abbreviated title of project

provided us with a completed questionnaire and gave us your name and address as a potential respondent from the community of practice. You were selected by the researcher as either the practitioner most involved in the applied research project indicated above or as being representative of those involved.

If you agree to participate, please take 20-30 minutes to fill out the attached questionnaire. Then simply fold your completed questionnaire and return it within two weeks of receipt in the self-addressed envelope (stamp provided). You will note that a code number has been written on the upper right hand corner of the questionnaire. The purpose of having this code number is to ensure that your responses are linked to those of the researcher listed above so that comparisons can be made. All responses will remain strictly **CONFIDENTIAL**. Only members of the immediate research team will have access individual questionnaire responses. These responses will be pooled with others for analysis. The master list of code numbers and names will be destroyed once all data have been entered onto the computer. Your name and address will not be used for any other purpose.

If you would like to receive a two-page summary of the survey results you can indicate this on the last page of the questionnaire. If you have inquiries about the project please contact the principal investigator at:

Mail Dr. Brad Cousins
Faculty of Education
University of Ottawa
145 Jean Jacques Lussier
Ottawa, ON, CANADA, K1N 6N5

Phone (613) 562-5800 Ext. 4088

Fax (613) 564-2475

E-mail bcousins@educ-1.edu.uottawa.ca

Thank you for your participation,

Brad Cousins
Associate Professor